Two-way classical communication remarkably improves local distinguishability

Masaki Owari [1,2] **and Masahito Hayashi** [3,4]

[1] Collaborative Institute of Nano Quantum Information Electronics, The University of Tokyo
Hongo 7-3-1, Bunkyo-ku Tokyo 113-0033, Japan

[2] Department of Physics, Graduate School of Science, The University of Tokyo
Hongo 7-3-1, Bunkyo-ku Tokyo 113-0033, Japan

[3] ERATO-SORST Quantum computation and information project, JST, Hongo 5-28-3, Bunkyo-ku Tokyo 113-0033, Japan

[4] Graduate School of Information Sciences, Tohoku University, Aoba-ku, Sendai, 980-8579, Japan

E-mail: masakiowari@is.s.u-tokyo.ac.jp, hayashi@math.is.tohoku.ac.jp

Abstract. We analyze the difference in the local distinguishability among the following three restrictions; (i) Local operations and only one-way classical communications (one-way LOCC) are permitted. (ii) Local operations and two-way classical communications (two-way LOCC) are permitted. (iii) All separable operations are permitted. We obtain two main results concerning the discrimination between a given bipartite pure state and the completely mixed state with the condition that the given state should be detected perfectly. As the first result, we derive the optimal discrimination protocol for a bipartite pure state in the cases (i) and (iii). As the second result, by constructing a concrete two-way local discrimination protocol, it is proven that the case (ii) is much better than the case (i), i.e., two-way classical communication remarkably improves the local distinguishability in comparison with one-way classical communication at least for a low-dimensional bipartite pure state.

PACS numbers: 03.65.Ud,03.65.Wj,03.67.Mn

Keywords: state discrimination, one-way LOCC, two-way LOCC, Schmidt rank, Schmidt coefficient

Submitted to: *New J. Phys.*

1. Introduction

Recently, quantum communication has been investigated among many groups as a future technology. Similar to conventional information technology, practical quantum communication technology will require distributed information processing among two or more spatially separated parties. In order to treat this problem, it is necessary to clarify what kind of information processing is possible under respective constraints for permitted operations. In the quantum case, when our quantum system consists of distinct two parties A and B, we often restrict our operations to local (quantum) operations and classical communications (LOCC) because sending quantum states over long distance is technologically more difficult than sending classical information[1]. Even in this restriction, we can consider the following two formulations; (i) The classical communication is restricted to the direction A to B (We can similarly treat the restriction of the opposite direction.) (ii) All parties are allowed to communicate classically with each other as much as they like. The case (i) is called the one-way LOCC, and the case (ii) is called the two-way LOCC.

Since, by definition, the two-way LOCC apparently includes the one-way LOCC, the two-way LOCC is always more powerful than the one-way LOCC in principle. However, due to the following two reasons, it is not easy to characterize the difference between the both performances. As the first reason, it is mathematically hard to rigorously evaluate the performance of a given two-way LOCC protocol, because the mathematical description of two-way LOCC is too complicated. As the second reason, for several simple tasks, the performance of the two-way LOCC was actually shown to be the same as the performance of the one-way LOCC. In fact, there are several settings that has no difference between one-way LOCC and two-way LOCC, e.g., LOCC convertibility of bipartite pure state [2], Stein's lemma bound in the simple asymptotic hypothesis testing of the n-tensor product of identical states[3].

On the other hand, in several settings, the two-way LOCC is strictly powerful than the one-way LOCC. For example, the distillable entanglement (the amount of maximally entangled states which can be derived from a given state by LOCC) with two-way classical communication is proven to be greater than that with one-way classical communication[4]. Also, this type comparison also has been done by several papers[5, 6, 7] only on the discrimination among orthogonal states. Although several researchers treated this problem, they did not treated the discrimination among non-orthogonal states. In this paper, we compare the both performances quantitatively on the "*local discrimination*" among states that are not necessarily orthogonal, whose purpose is discriminating given states by only LOCC with the *single* copy. In fact, general discrimination problem is closely related to sending classical information via quantum channel[8] and quantum algorithm [9, 10].

In order to quantify the difference of the two-way LOCC and the one-way LOCC, in this paper, we concentrate on a simple setting: the local discrimination of the first state ρ on a bipartite system $\mathcal{H}$ from the second state $\tilde{\rho}$ under the condition where

the first state ρ should be detected perfectly. When the both states ρ and $\tilde{\rho}$ are pure, there is no difference between one-way LOCC and two-way LOCC because any global discrimination protocol can be simulated by one-way LOCC[11, 12]. Surprisingly, as our result, we found that there usually exists non-negligible difference between two restrictions when the second $\tilde{\rho}$ is the completely mixed state $\rho_{\mathrm{mix}} := I/\dim\mathcal{H}$. At the first glance, this setting seems specific, however, due to the following six reasons, it is closely related to several research topics. First, this type analysis produces a bound of the number of perfectly locally distinguishable states. Second, as is explained later, there is a relation between the performance of local distinguishability and amount of entanglement in the case of pure states. Third, this kind of distinguishability is often treated in quantum complexity as Triviality of Coset State [10, 13]. Fourth, when the second state $\tilde{\rho}$ is close to the completely mixed state ρ_{mix}, we obtain a similar conclusion because the power of our test is continuous concerning the second state. Fifth, in the community of classical statistics, the problem of discriminating the given two distributions is widely accepted as the fundamental problem of hypothesis testing because general hypothesis testing problem can be treated by using this type problem[14]. Sixth, as was mentioned in the preceding papers [8], hypothesis testing with two candidates states is closely related to quantum channel coding. Hence, it is suitable to treat this kind of local discrimination problem.

In order to analyze this problem in the respective settings, we introduce the minimum error probabilities to detect the complete mixed state $\beta_{\rightarrow}(\rho)$, $\beta_{\leftrightarrow}(\rho)$, and $\beta_{\mathrm{sep}}(\rho)$ by the one-way LOCC, the two-way LOCC, and the separable operations, respectively. Indeed, these functions are considered as appropriate measures of the local distinguishability because they give not only the minimum error probability of the above problem, but also the upper bound of the size of locally distinguishable sets in general perfect local discrimination problems[15]. Under this formulation, we first analyze the local distinguishability by means of one-way LOCC and separable operations, and derive the optimal discrimination protocol with one-way LOCC and separable operations; we should note that the minimum error probability $\beta_{\mathrm{sep}}(\rho)$ with separable operations gives a lower bound for the minimum error probability $\beta_{\leftrightarrow}(\rho)$ with two-way LOCC. After that, constructing a concrete two-way local discrimination protocol, we show that two-way classical communication remarkably improves the local distinguishability in comparison with the local discrimination by one-way classical communication at least for a low- (less than five) dimensional bipartite pure state. Indeed, since the power of our test is continuous concerning the first and the second states, our result indicates that two-way classical communication remarkably improves the local distinguishability in a wider class of the first and the second states. Moreover, as a byproduct, we extend the characterization of locally distinguishability by one-way LOCC by Cohen [7] to a set of mixed states.

This paper is organized as follows: In Section 2, we introduce the discrimination problem between an arbitrary given state ρ and a completely mixed state ρ_{mix} on a bipartite system $\mathcal{H}$ under the condition that the given state is detected perfectly. Then,

we explain another meaning of $\beta_{\rightarrow}(\rho)$, $\beta_{\leftrightarrow}(\rho)$, and $\beta_{\mathrm{sep}}(\rho)$ from the viewpoint of general local discrimination problems. In Section 3, constructing the optimal separable POVM for the local discrimination, we prove that $D\beta_{\mathrm{sep}}(|\Psi\rangle)-1$ coincides with the entanglement monotone called robustness of the entanglement for a bipartite pure state, where D is the dimension of the bipartite Hilbert space $\mathcal{H}$. In Section 4, we show that the amount $D\beta_{\rightarrow}(|\Psi\rangle)$ with one-way LOCC coincides with the Schmidt rank (the rank of the reduced density matrix) of the states. Also, as a corollary, we extend Cohen's characterization to a set of mixed states. Finally, in section 5, constructing a concrete three-step two-way LOCC discrimination protocol, we derive an upper bound for $\beta_{\leftrightarrow}(\rho)$. Calculating this upper bound analytically and also numerically, we show that $\beta_{\leftrightarrow}(|\Psi\rangle)$ is strictly smaller than $\beta_{\rightarrow}(|\Psi\rangle)$, and moreover, $\beta_{\rightarrow}(\rho)$ and $\beta_{\mathrm{sep}}(\rho)$ give almost the same value for a lower dimensional bipartite pure state; this results can be seen in **Figures 2,3,4,5,6**. As a result, we conclude that the two-way classical communication remarkably improves the local distinguishability in comparison with the one-way classical communication for a low-dimensional pure state at least in the present problem settings.

2. Local discrimination between an arbitrary state and the completely mixed state

In this paper, we treat the bipartite system $\mathcal{H} := \mathcal{H}_A \otimes \mathcal{H}_B$ ($\dim \mathcal{H} = D$) composed of two finite-dimensional subsystems $\mathcal{H}_A$ and $\mathcal{H}_B$. In the following sections, we often focus on the case when ρ is pure. In such a case, we assume that the dimension d of $\mathcal{H}_A$ is equal to that of $\mathcal{H}_B$. Note that the given pure state belongs to the composite system of the same-dimensional subsystem. Then, the dimension (D) of the Hilbert space $\mathcal{H}$ is equal to d^2. In the composite system $\mathcal{H}$, we call a positive operator T with $0 \le T \le I$ a one-way LOCC POVM element, where I is an identity operator on $\mathcal{H}$, if the two-valued POVM $\{T, I-T\}$ can be implemented by the one-way LOCC; we also define a two-way LOCC POVM element and a separable POVM element in the same manner by using the two-way LOCC and the separable operations in stead of the one-way LOCC, respectively [16]. We write a set of one-way LOCC, two-way LOCC, separable POVM elements, and all (global) POVM elements as $\mathcal{T}_{\rightarrow}$, $\mathcal{T}_{\leftrightarrow}$, $\mathcal{T}_{\mathrm{sep}}$, and $\mathcal{T}_{\mathrm{g}}$. Obviously, they satisfy the relation $\mathcal{T}_{\rightarrow} \subset \mathcal{T}_{\leftrightarrow} \subset \mathcal{T}_{\mathrm{sep}} \subset \mathcal{T}_{\mathrm{g}}$. We can see that the condition $T \in \mathcal{T}_c$ is equivalent with the condition $I - T \in \mathcal{T}_c$, where c can be either $\rightarrow, \leftrightarrow, \mathrm{sep}$, or g.

In this paper, we discuss the comparison of the performance of the local discrimination in the case of the one-way LOCC, the two-way LOCC, and the separable operations. In order to find this difference, although there are many problem settings for the local discrimination, we especially focus on one of the simplest problem settings as follows: We consider local discrimination of an given arbitrary state ρ and another state $\tilde{\rho}$, and investigate how well we can detect $\tilde{\rho}$ under the additional condition that we do not make any error to detect ρ when the second state $\tilde{\rho}$ is the completely mixed state $\rho_{\mathrm{mix}} \stackrel{\mathrm{def}}{=} \frac{I_{AB}}{D}$ ($D = \dim \mathcal{H}$); namely, by only LOCC, how well we can distinguish a given entangled state ρ from the white noise state ρ_{mix} without making any error to

judge the given state is $\rho_{\rm mix}$ when the real state is ρ.

Our problem can be written down rigorously as follows. We measure an unknown state chosen from two candidates $\{\rho, \tilde{\rho}\}$ by the two-values POVM $\{T, I-T\}$, where $T \in \mathcal{T}_{\rightarrow}, \mathcal{T}_{\leftrightarrow}, \mathcal{T}_{\rm sep}$, or $\mathcal{T}_{\rm g}$; that is, if we get the result corresponding to T, then we decide that the unknown state is in ρ, and if we get the result corresponding to $I-T$, then we decide that the unknown state is in $\tilde{\rho}$. We consider two kinds of error probability as follows: the type 1 error probability $\operatorname{Tr}\rho(I-T)$, and the type 2 error probability $\operatorname{Tr}\tilde{\rho}T$; these are common terms in the field of "*quantum hypothesis testing*" [3], where these two different error probabilities are treated in an asymmetric way. In this case, the type 1 error probability corresponds to the error probability that the real state is ρ and our decision is $\tilde{\rho}$, and the type 2 error probability corresponds to the error probability that the real state is $\tilde{\rho}$ and our decision is ρ. Thus, our problem is to minimize the type 2 error probability $\operatorname{Tr}\tilde{\rho}T$ under the additional condition that the type 1 error probability $\operatorname{Tr}\rho(I-T)$ must be 0. Thus, we focus on the following minimum of the type 2 error probability:

$$\beta_c(\rho\|\tilde{\rho}) := \min\{\operatorname{Tr}(\tilde{\rho}T) | T \in \mathcal{T}_c, \operatorname{Tr}\rho T = 1\}, \tag{1}$$

where $c = \rightarrow$(one-way LOCC), $\leftrightarrow$(two-way LOCC), sep(separable operations), and g(global operations). When the both states ρ and $\tilde{\rho}$ are pure states $|\Phi\rangle$ and $|\Psi\rangle$, this quantity does not depend on whether $c = \rightarrow, \leftrightarrow, \text{sep}, \text{or g}$, and is calculated as

$$\beta_c(|\Phi\rangle \| |\Psi\rangle) = |\langle\Phi|\Psi\rangle|^2 \tag{2}$$

for $c = \rightarrow$, $\leftrightarrow$, sep, and g. This is because any discriminating protocol between two pure bipartite states can be simulated by one-way LOCC when we focus only on the distribution of the outcome[11, 12]. In this paper, we focus on the minimum of the type 2 error probability in the case of $\tilde{\rho} = \rho_{\rm mix}$:

$$\beta_c(\rho) := \beta_c(\rho\|\rho_{\rm mix}) = \frac{t_c(\rho)}{D}, \tag{3}$$

where $t_c(\rho)$ is defined as

$$t_c(\rho) = \min\{\operatorname{Tr} T | T \in \mathcal{T}_c, \operatorname{Tr} T\rho = 1\}. \tag{4}$$

and D is the dimension of the whole system $\mathcal{H}$. That is, $t_c(\rho)$ is in proportion to the minimum of the type 2 error probability $\beta_c(\rho)$ of one-way LOCC, two-way LOCC, separable POVM and global POVM in the case where $c = \rightarrow$, $\leftrightarrow$, sep, and g, respectively. Trivially,

$$t_{\rm g}(\rho) = \operatorname{rank}\rho. \tag{5}$$

Obviously, $t_c(\rho)$ satisfies the inequality $t_{\rm g}(\rho) \le t_{\rm sep}(\rho) \le t_{\leftrightarrow}(\rho) \le t_{\rightarrow}(\rho)$; as a matter of course, $\beta_c(\rho)$ also satisfies the similar inequality. Note that by normalizing $\beta_c(\rho)$ as the above Eq.(3), the resulting function $t_c(\rho)$ is no more a function depending both on ρ and $\rho_{\rm mix}$, but a function depending only on ρ.

Remark 1 In quantum information community, many papers treats the Bayesian framework, in which the Bayesian prior distribution is assumed [17, 18, 19]. However, in statistics community, non-Bayesian framework is more widely accepted, in which no Bayesian prior distribution is assumed[14]. This is because it is usually quite difficult to decide the Bayesian prior distribution based on the prior knowledge. In order to resolve this difficulty, they often treat the two kinds of error probabilities in an asymmetric way in hypothesis testing without assuming prior distribution because the importance of both error are not equal in a usual case, e.g., Neyman-Pearson lemma[14], Stein's lemma[20], Hoeffding bound[21]. These quantum cases are treated by several papers[22, 23, 24, 25]. In this paper, according to conventional statistics framework, we focus on the error probabilities of the first and second, and minimize the second kind of error probability under the constraint for the first one.

Here, we explain the reason why we choose the above special problem of discrimination of an arbitrary state ρ from the completely mixed state $\rho_{\rm mix}$, and the reason why we add the above additional condition of perfect detection of ρ. As we already said before, the first reason is that this additional condition makes the analysis of the problem extremely easier. Actually as we will see later in this paper, we can derive the optimal POVM of this restricted local-discrimination problem with respect to each one-way LOCC and separable operations for a bipartite pure state. As a result, we make the difference between one-way LOCC and two-way LOCC clear for our local-discrimination problem; this is our main purpose in this paper. Note that it is generally a hard problem to find an optimal protocol for a local-discrimination problem, and only in very limited situations, optimal local-discrimination protocols are known [5, 11, 12]. The second reason is that we can clearly see the relationship between local distinguishability and entanglement of a state in this problem setting. In the previous paper [15], we showed the relationship between local distinguishability of a set of states and an average of the values of entanglement monotones for the states in terms of inequalities. However, in this paper, we will show that the minimum error probability $\beta_c(\rho)$ of our problem is proportion to entanglement monotones in the case of one-way LOCC ($c =\rightarrow$) and separable operations ($c = \mathrm{sep}$) at least for bipartite pure states except an unimportant constant factor. The third reason is that the minimum error probability $\beta_c(\rho)$ can give a bound of local distinguishability for a more general local discrimination problem: Suppose that a set of states $\{\rho_i\}_{i=1}^{N_c}$ is perfectly locally distinguishable by one-way LOCC ($c =\rightarrow$), two-way LOCC ($c =\leftrightarrow$), or separable ($c = \mathrm{sep}$) POVM. From the result obtain in the the previous paper, $t_c(\rho_i)$ (which corresponds to $d(\rho)$) gives an upper bound of N_c as [15],

$$N_c \leq D/\overline{t_c(\rho_i)} = 1/\overline{\beta_c(\rho_i)}, \tag{6}$$

where $\overline{t_c(\rho_i)}$ and $\overline{\beta_c(\rho_i)}$ are the average of $\{t_c(\rho_i)\}_{i=1}^{N_c}$ and $\{\beta_c(\rho_i)\}_{i=1}^{N_c}$, respectively [15]. Thus, $\beta_c(\rho)$ can be considered as an appropriate measure of local distinguishability in a original operational sense, and also as a function whose average gives an upper bound for the locally distinguishable sets of states. Therefore, we investigate the difference

of local distinguishability of ρ by one-way LOCC POVM, two-way LOCC POVM, and separable POVM in terms of $\beta_c(\rho)$ in the following sections.

3. Local discrimination by separable POVM

In this section, we investigate the minimum type 2 error probability $\beta_{\rm sep}(\rho) = \frac{t_{\rm sep}(\rho)}{D}$ in terms of separable POVMs, which are given by $\{N_i \otimes M_i\}_i$ with the conditions $\sum_i N_i \otimes M_i = I$, $N_i \geq 0$, and $M_i \geq 0$. The main purpose of this section is proving the following theorem:

Theorem 1 The inequality

$$t_{\rm sep}(\rho) \geq \max\{(\mathrm{Tr}\sqrt{\rho_A})^2, (\mathrm{Tr}\sqrt{\rho_B})^2\} \tag{7}$$

holds for a bipartite state ρ on $\mathcal{H}_A \otimes \mathcal{H}_B$, where ρ_A and ρ_B are the reduced density matrix of $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Any pure state satisfies its equality. In other words, the following inequality concerning the minimum error probability $\beta_{\rm sep}(\rho)$ holds:

$$\beta_{\rm sep}(\rho) \geq \frac{1}{D}\max\{(\mathrm{Tr}\sqrt{\rho_A})^2, (\mathrm{Tr}\sqrt{\rho_B})^2\}. \tag{8}$$

For a bipartite pure state, the right-hand side of Eq.(7) is proportional to an entanglement monotone called the global robustness of entanglement $R_g(|\Psi\rangle)$ except an unimportant constant term [26].

Applying Theorem 1 for Eq.(6), we can immediately derive the following corollary concerning the perfect discrimination of a given set of states in term of separable operations:

Corollary 1 If a set of states $\{\rho_i\}_{i=1}^N$ is perfectly distinguishable by separable operations, then, the set of states $\{\rho_i\}_{i=1}^N$ satisfies the following inequality:

$$N \leq D/\overline{\max\{(\mathrm{Tr}\sqrt{\rho_{iA}})^2, (\mathrm{Tr}\sqrt{\rho_{iB}})^2\}}, \tag{9}$$

where $\overline{\max\{(\mathrm{Tr}\sqrt{\rho_{iA}})^2, (\mathrm{Tr}\sqrt{\rho_{iB}})^2\}}$ is the average of $\max\{(\mathrm{Tr}\sqrt{\rho_{iA}})^2, (\mathrm{Tr}\sqrt{\rho_{iB}})^2\}$ for all $1 \leq i \leq N$.

The above inequality is weaker than the inequality (6). However, the inequality (9) is superior to the inequality (6) in terms of the efficiency of the computation; that is, in general, we can not efficiently compute the bound in Eq.(6), since the function $t_{\rm sep}(\rho)$ includes the big variational problem.

3.1. Pure states case

First, for a technical reason, we concentrate the pure states case, and define a set of POVM elements $\mathcal{T}_{\widetilde{\rm sep}}$ by,

$$\mathcal{T}_{\widetilde{\rm sep}} \stackrel{\rm def}{=} \left\{ T \;\middle|\; T \leq I_{AB}, T = \sum_i N_i \otimes M_i, \quad \forall i, N_i \geq 0, M_i \geq 0 \right\}.$$

$\mathcal{T}_{\widetilde{\mathrm{sep}}}$ is a set of POVM elements can be decomposed into a separable form; we say a positive linear operator M has a separable form, if $M/\operatorname{Tr} M$ is a separable state. Since the definition of $\mathcal{T}_{\widetilde{\mathrm{sep}}}$ is equivalent to the definition of $\mathcal{T}_{\mathrm{sep}}$ except the condition $I-T \in \mathcal{T}_{\mathrm{sep}}$, $\mathcal{T}_{\mathrm{sep}}$ is a subset of $\mathcal{T}_{\widetilde{\mathrm{sep}}}$. Note that even if $T \in \mathcal{T}_{\widetilde{\mathrm{sep}}}$, $I-T$ does not necessary satisfy $I-T \in \mathcal{T}_{\widetilde{\mathrm{sep}}}$; that is, $\mathcal{T}_{\mathrm{sep}}$ does not coincide with $\mathcal{T}_{\widetilde{\mathrm{sep}}}$. For example, suppose a set of states $\{|\Psi_i\rangle\}_{i=1}^m \subset \mathcal{H}_A \otimes \mathcal{H}_B$ ($m < \dim \mathcal{H}$) is an unextendable product basis, and a POVM T is defined as $T \stackrel{\text{def}}{=} \sum_i |\Psi_i\rangle\langle\Psi_i|$. Then, T belongs to $\mathcal{T}_{\widetilde{\mathrm{sep}}}$, but not to $\mathcal{T}_{\mathrm{sep}}$ since $I - T = I - \sum_{i=1}^m |\Psi_i\rangle\langle\Psi_i|$ is proportion to a (bound) entangled state, and does not a separable form [27]. Similarly, we can define $t_{\widetilde{\mathrm{sep}}}$ as,

$$t_{\widetilde{\mathrm{sep}}}(\rho) \stackrel{\text{def}}{=} \min \left\{ \operatorname{Tr} T | T \in \mathcal{T}_{\widetilde{\mathrm{sep}}}, \operatorname{Tr} \rho T = 1 \right\}. \tag{10}$$

By definition, $t_{\widetilde{\mathrm{sep}}}(\rho)$ apparently gives a lower bound of $t_{\mathrm{sep}}(\rho) = d^2 \beta_{\mathrm{sep}}(\rho)$, that is, for all $\rho \in S(\mathcal{H})$,

$$t_{\widetilde{\mathrm{sep}}}(\rho) \leq t_{\mathrm{sep}}(\rho) = d^2 \beta_{\mathrm{sep}}(\rho). \tag{11}$$

Then, we can see that $t_{\widetilde{\mathrm{sep}}}(\rho)$ is actually equal to $d(\rho)$ which is defined in Theorem 1 of the paper [15] as:

$$d(\rho) \stackrel{\text{def}}{=} \min \left\{ \frac{1}{\operatorname{Tr} \rho\omega} \middle| 0 \leq \frac{\omega}{\operatorname{Tr} \rho\omega} \leq I, \omega \in \mathrm{SEP} \right\}, \tag{12}$$

where SEP is the set of all separable states. We can easily check this fact just by defining $T \stackrel{\text{def}}{=} \frac{\omega}{\operatorname{Tr} \rho\omega}$; then, T satisfies $0 \leq T \leq I$, $\operatorname{Tr} \rho T = 1$, and $T \in \mathcal{T}_{\widetilde{\mathrm{sep}}}$. From Theorem 2 of the paper [15], for an arbitrary multipartite pure state $|\Psi\rangle$, $t_{\widetilde{\mathrm{sep}}}(\rho)$ satisfies the following inequality:

$$t_{\widetilde{\mathrm{sep}}}(|\Psi\rangle) = d(|\Psi\rangle) \geq 1 + R_g(|\Psi\rangle), \tag{13}$$

and $R_g(|\Psi\rangle)$ is the global robustness of entanglement [26] defined as:

$$R_g(\rho) \stackrel{\text{def}}{=} \min \left\{ t \geq 0 \middle| \exists \text{a state } \Delta, \text{ s.t. } \frac{1}{1+t}(\rho + t\Delta) \in \mathrm{SEP} \right\}. \tag{14}$$

For a bipartite pure state, we can know a more detail of $t_{\widetilde{\mathrm{sep}}}(|\Psi\rangle)$ as follows. First, it was proven that $t_{\widetilde{\mathrm{sep}}}(|\Psi\rangle)$ coincides with the robustness of entanglement $R_g(|\Psi\rangle)$ for an arbitrary pure bipartite state $|\Psi\rangle$ [28]. This fact can be seen by checking that the optimal states of $R_g(\rho)$, which was derived in [26] satisfies the condition of $d(\rho)$; the optimal state of $R_g(\rho)$ is also an optimal state of $d(\rho)$. Moreover, we know that the value of $R_g(|\Psi\rangle)$ is given by the following formula for a bipartite state $|\Psi\rangle$ [26]:

$$R_g(\rho) = \Big(\sum_i \sqrt{\lambda_i} \Big)^2 - 1,$$

where $\{\lambda_i\}_{i=1}^d$ is the Schmidt coefficients of $|\Psi\rangle$; $|\Psi\rangle$ can be decomposed as $|\Psi\rangle = \sum_i \sqrt{\lambda_i} |e_i\rangle \otimes |f_i\rangle$ by choosing an appropriate orthonormal basis sets of the local Hilbert spaces $\{|e_i\rangle\}_{i=1}^d \subset \mathcal{H}_A$ and $\{|f_i\rangle\}_{i=1}^d \subset \mathcal{H}_B$. Thus, we derive

Lemma 1 For a bipartite pure state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$,

$$t_{\widetilde{\mathrm{sep}}}(|\Psi\rangle) = d(|\Psi\rangle) = 1 + R_g(|\Psi\rangle) = \Big(\sum_i \sqrt{\lambda_i}\Big)^2. \tag{15}$$

Although this lemma is a known result [28], as a preparation of the proof of the next theorem, we give a complete proof of Eq.(15), in which we prove directly the equation $t_{\widetilde{\mathrm{sep}}}(|\Psi\rangle) = \left(\sum_i \sqrt{\lambda_i}\right)^2$ from the definition of $t_{\widetilde{\mathrm{sep}}}(|\Psi\rangle)$.

Proof This proof is divided into two-steps. In the first step, we prove that $\left(\sum_i \sqrt{\lambda_i}\right)^2$ is the lower bound of $t_{\widetilde{\mathrm{sep}}}(|\Psi\rangle)$. Then, in the second step, we construct POVM element T which attains this lower bound. For convenience, we define $|M_\Psi\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^d |e_i\rangle \otimes |f_i\rangle$ where $\{|e_i\rangle\}_{i=1}^d$ and $\{|f_i\rangle\}_{i=1}^d$ are the Schmidt basis of $|\Psi\rangle$; thus, $|M_\Psi\rangle$ is the maximum entangled state sharing the Schmidt basis with $|\Psi\rangle$. Then, we derive $d|\langle M_\Psi|\Psi\rangle|^2 = \left(\sum_i \sqrt{\lambda_i}\right)^2$.

As the first step, we prove the following inequality;

$$\begin{aligned} t_{\widetilde{\mathrm{sep}}}(|\Psi\rangle) = &\ \min\{\mathrm{Tr}\, T | 0 \le T \le I,\ T \text{ is sep},\ \langle\Psi| T |\Psi\rangle = 1\} \\ \ge &\ \min\{d\,\langle M_\Psi| T |M_\Psi\rangle\,| 0 \le T \le I, T \text{ is sep}, \langle\Psi| T |\Psi\rangle = 1\} \\ \ge &\ d|\langle M_\Psi|\Psi\rangle|^2. \end{aligned} \tag{16}$$

To prove the first inequality (16), since both $\mathrm{Tr}\, T$ and $d\,\langle M_\Psi| T |M_\Psi\rangle$ are linear for T, it is enough to prove only in the case that T can be written down as $T = |a\rangle\langle a| \otimes |b\rangle\langle b|$ by using un-normalized vectors $|a\rangle$ and $|b\rangle$. Suppose $|a\rangle = \sum_{i=1}^d \alpha_i |e_i\rangle$ and $|b\rangle = \sum_{i=1}^d \beta_i |f_i\rangle$. Then, using Schwarz's inequality, we can prove as follows

$$\mathrm{Tr}\, T = (\sum_i |\alpha_i|^2)(\sum_i |\beta_i|^2) \ge |\sum_j \alpha_j \beta_j|^2 = d\,\langle M_\Psi| T |M_\Psi\rangle\,.$$

For the second inequality (16), since the relations $\langle\Psi| T |\Psi\rangle = 1$ and $T \le I$ deduce that $|\Psi\rangle$ is an eigenvector of the largest eigenvalue 1 of T, we derive $T \ge |\Psi\rangle\langle\Psi|$. Therefore, the inequality $d\,\langle M_\Psi| T |M_\Psi\rangle \ge d|\langle M_\Psi|\Psi\rangle|^2$ is derived by taking the mean value with respect to $|M_\Psi\rangle$.

As the second step, we construct an example of POVM element T which achieves the lower bound we derived above. Define T_0 as $T_0 = |a\rangle\langle a| \otimes |b\rangle\langle b|$ where $|a\rangle = \sum_{i=1}^d (\lambda_i)^{1/4} |e_i\rangle$ and $|b\rangle = \sum_{i=1}^d (\lambda_i)^{1/4} |f_i\rangle$; then, T_0 satisfies $\mathrm{Tr}\, T_0 = \left(\sum_i \sqrt{\lambda_i}\right)^2$. Moreover, since $P_0 T_0 P_0 = |\Psi\rangle\langle\Psi|$ where $P_0 = \sum_{i=1}^d |e_i\rangle\langle e_i| \otimes |f_i\rangle\langle f_i|$, T_0 satisfies $\langle\Psi| T_0 |\Psi\rangle = 1$. Since T_0 apparently satisfies $0 \le T_0$, the inequality $T \le I$ is the only remaining condition which the optimal POVM element T attaining the equality $\mathrm{Tr}\, T = t_{\widetilde{\mathrm{sep}}}(|\Psi\rangle)$ must satisfy. Since T_0 does not generally satisfies the inequality $T_0 \le I$, we construct a new POVM element T which satisfies $0 \le T \le I$ from T_0. In order to construct the POVM element T from T_0, we use twirling technique here. We define a

family of local unitary operators $U_{\overrightarrow{\theta}}$ parameterized by $\overrightarrow{\theta} = \{\theta_i\}_{i=1}^d$ as follows,

$$U_{\overrightarrow{\theta}} = (\sum_{j=1}^{d} e^{i\theta_j} |e_j\rangle \langle e_j|) \otimes (\sum_{k=1}^{d} e^{-i\theta_k} |f_k\rangle \langle f_k|). \tag{17}$$

Note that $\left(\mathcal{H}^{\otimes 2}, U_{\overrightarrow{\theta}}\right)$ is a unitary representation of the compact topological group $\overbrace{U(1) \times \cdots \times U(1)}^{n}$; by means of a unitary representation of a compact topological group, we implement the "*twirling*" operation (the averaging over the compact topological group) for a state (or POVM) [29]. Then, we define T as the operator which is constructed by twirling $U_{\overrightarrow{\theta}} T_0 U_{\overrightarrow{\theta}}^{\dagger}$ over parameters $\overrightarrow{\theta} = \{\theta_i\}_{i=1}^d$. Since by an action of twirling operation, a given state is projected to the subspace of all invariant elements of the group action [29], we can calculate T as follows:

$$\begin{aligned} T &\stackrel{\text{def}}{=} \int_0^{2\pi} \cdots \int_0^{2\pi} U_{\overrightarrow{\theta}} T_0 U_{\overrightarrow{\theta}}^{\dagger} d\theta_1 \cdots d\theta_d \\ &= (\sum_{j=1}^{d} |e_j\rangle \langle e_j| \otimes |f_j\rangle \langle f_j|) T_0 (\sum_{j=1}^{d} |e_j\rangle |e_j\rangle \otimes |f_j\rangle \langle f_j|) \\ &\qquad + \sum_{j \neq k} (|e_j\rangle \langle e_j| \otimes |f_k\rangle \langle f_k|) \, T_0 \, (|e_j\rangle \langle e_j| \otimes |f_k\rangle \langle f_k|) \\ &= (\sum_{i=1}^{d} \sqrt{\lambda_i} \, |e_i\rangle \otimes |f_i\rangle)(\sum_{i=1}^{d} \sqrt{\lambda_i} \, \langle e_i| \otimes \langle f_i|) \\ &\qquad + \sum_{i \neq j} \sqrt{\lambda_i \lambda_j} \, |e_i\rangle \langle e_i| \otimes |f_j\rangle \langle f_j| \, . \end{aligned}$$

Since $\sqrt{\lambda_i \lambda_j} \leq 1$, $T \leq I$. Moreover, T satisfies $0 \leq T \leq I$, $\langle \Psi | T | \Psi \rangle = 1$, and is in the separable form; we only applied the local unitary $U_{\overrightarrow{\theta}}$ to un-normalized product state T_0, and, then, took an average over parameters $\overrightarrow{\theta}$. Thus, we derive the inequality $t_{\widetilde{\text{sep}}}(|\Psi\rangle) \leq \operatorname{Tr} T = \left(\sum_i \sqrt{\lambda_i}\right)^2$. Since we have already proven the converse inequality, we conclude $t_{\widetilde{\text{sep}}}(|\Psi\rangle) = \left(\sum_i \sqrt{\lambda_i}\right)^2$
□

Finally, by means of Lemma 1, we can derive the following theorem, i.e., Theorem 1 in the pure states case:

Theorem 2 For a bipartite pure state $|\Psi\rangle$,

$$\begin{aligned} \beta_{\text{sep}}(|\Psi\rangle) = \frac{1}{d^2} t_{\text{sep}}(|\Psi\rangle) &= \frac{1}{d^2} \left(1 + R_g(|\Psi\rangle)\right) \\ &= \frac{1}{d^2} \Big(\sum_i \sqrt{\lambda_i}\Big)^2 = \frac{1}{d^2} (\operatorname{Tr} \sqrt{\rho_A})^2 = \frac{1}{d^2} (\operatorname{Tr} \sqrt{\rho_B})^2, \end{aligned} \tag{18}$$

where $\{\lambda_i\}_{i=1}^d$ is the Schmidt coefficients of $|\Psi\rangle$.

Proof Since by the definition $t_{\widetilde{\text{sep}}}(|\Psi\rangle) \leq t_{\text{sep}}(|\Psi\rangle)$, all what we need to prove is that the optimal POVM T for $t_{\widetilde{\text{sep}}}(|\Psi\rangle)$ is also the optimal POVM for $t_{\text{sep}}(|\Psi\rangle)$; that is, $I-T$ also has a separable form.

As we have already shown in the proof of Lemma 1, the optimal POVM T for $t_{\widetilde{\text{sep}}}(|\Psi\rangle)$ can be written down as

$$T = (\sum_{i=1}^{d} \sqrt{\lambda_i}\, |e_i\rangle \otimes |f_i\rangle)(\sum_{i=1}^{d} \sqrt{\lambda_i}\, \langle e_i| \otimes \langle f_i|) \, + \sum_{i \neq j} \sqrt{\lambda_i \lambda_j}\, |e_i\rangle \langle e_i| \otimes |f_j\rangle \langle f_j| \,, \tag{19}$$

where $\{|e_i\rangle \otimes |f_j\rangle\}_{ij}$ and $\{\lambda_i\}_{i=1}^{d}$ are the Schmidt basis and the Schmidt coefficients corresponding to $|\Psi\rangle = \sum_{i=1}^{d} \sqrt{\lambda_i}\, |e_i f_j\rangle$, respectively. Suppose

$$\overline{T_0} \stackrel{\text{def}}{=} \frac{1}{2} \sum_{i \neq j} |\overline{a_{ij}}\rangle \langle \overline{a_{ij}}| \otimes |\overline{b_{ij}}\rangle \langle \overline{b_{ij}}| + \sum_{i \neq j} \{ \sum_{k \neq i,j} \lambda_k + (\sqrt{\lambda_i} - \sqrt{\lambda_j})^2 \} \, |e_i f_j\rangle \langle e_i f_j| \,,$$

where $|\overline{a_{ij}}\rangle$ and $|\overline{b_{ij}}\rangle$ are defined as $|\overline{a_{ij}}\rangle \stackrel{\text{def}}{=} (\lambda_j)^{\frac{1}{4}} |e_i\rangle - (\lambda_i)^{\frac{1}{4}} |e_j\rangle$ and $|\overline{b_{ij}}\rangle \stackrel{\text{def}}{=} (\lambda_j)^{\frac{1}{4}} |f_i\rangle + (\lambda_i)^{\frac{1}{4}} |f_j\rangle$ for $i \neq j$, respectively. Then, as is proven in Appendix A, the relation

$$\int_0^{2\pi} \cdots \int_0^{2\pi} U_{\overrightarrow{\theta}} \overline{T_0} U_{\overrightarrow{\theta}}^{\dagger} d\theta_1 \cdots d\theta_d = I - T \tag{20}$$

holds, where a local unitary operator $U_{\overrightarrow{\theta}}$ is defined as Eq.(17). By the definition, $\int_0^{2\pi} \cdots \int_0^{2\pi} U_{\overrightarrow{\theta}} \overline{T_0} U_{\overrightarrow{\theta}}^{\dagger} d\theta_1 \cdots d\theta_d$ is apparently a separable POVM element. Therefore, we can conclude the equality $t_{\widetilde{\text{sep}}}(|\Psi\rangle) = t_{\text{sep}}(|\Psi\rangle)$ for a bipartite pure state. By means of Lemma 1, we derive Eq.(18). □

Thus, $t_{\text{sep}}(|\Psi\rangle)$ is equivalent to $1 + R_g(|\Psi\rangle)$, and the minimum type 2 error probability $\beta_{\text{sep}}(|\Psi\rangle)$ only depends on the global robustness of entanglement $R_g(|\Psi\rangle)$ for a bipartite pure state $|\Psi\rangle$. In this case, the optimal POVM $\{T, I-T\}$ can be derived by using Eq.(19) as the definition of the POVM element T.

We should note that Theorem 2 not only gives a way to calculate the minimum type 2 error probability under separable operations $\beta_{\text{sep}}(|\Psi\rangle)$, but this theorem gives a complete relationship between the local distinguishability of a bipartite state under separable operations and the entanglement of the state. In the previous paper [15], it was shown the global robustness of entanglement $R_g(|\Psi\rangle)$ gives an upper-bound for the maximum number of distinguishable states under separable operations. However, the present result shows that $R_g(|\Psi\rangle)$ is nothing but the local distinguishability (against the completely mixed state) itself at least for a bipartite pure state. In other words, it is shown that robustness of entanglement has rigorously operational meaning for bipartite pure states in terms of the local discrimination from the completely mixed state ρ_{mix}.

3.2. Mixed states case

Now, we prove Theorem 1 for a general mixed bipartite state.

Proof (Theorem 1) First we prove the inequality $t_{\text{sep}}(\rho) \geq (\text{Tr}\,\sqrt{\rho_A})^2$. Adding the system B', we choose a purification $|\Phi\rangle$ of ρ. In the following, we will prove the inequality

$t_{\widetilde{\rm sep}}(\rho) \ge t_{\widetilde{\rm sep}}(|\Phi\rangle)$. If this inequality holds, applying Eq.(11) and Eq.(15), we obtain $t_{\rm sep}(\rho) \ge t_{\widetilde{\rm sep}}(\rho) \ge t_{\widetilde{\rm sep}}(|\Phi\rangle) = (\mathrm{Tr}\sqrt{\rho_A})^2$.

Define a separable positive operator $T = \sum_i p_i|e_i\rangle\langle e_i| \otimes |f_i\rangle\langle f_i|$ ($e_i \in \mathcal{H}_A, f_i \in \mathcal{H}_B, \|f_i\| = 1, \|e_i\| = 1$) such that $0 \le T \le I_{AB}$ and $\mathrm{Tr}\,\rho T = 1$. Thus, $\langle\Phi| \sum_i p_i(|e_i\rangle\langle e_i| \otimes |f_i\rangle\langle f_i| \otimes I_{B'})|\Phi\rangle = 1$. Now, we focus on the bipartite system A and BB'. Then, we choose the state $|\tilde{f}_i\rangle$ ($\|\tilde{f}_i\| = 1$) on $\mathcal{H}_{BB'}$ such that $\mathrm{Tr}_A(|\Phi\rangle\langle\Phi|)(|e_i\rangle\langle e_i| \otimes I_{BB'}) = c_i|\tilde{f}_i\rangle\langle\tilde{f}_i|$, where c_i is the normalizing constant. Define the state $|f'_i\rangle$ ($\|f'_i\| = 1$) on $\mathcal{H}_{BB'}$ by

$$|f'_i\rangle\langle f'_i| := \frac{1}{\langle\tilde{f}_i|P_i|\tilde{f}_i\rangle} P_i|\tilde{f}_i\rangle\langle\tilde{f}_i|P_i \le P_i, \tag{21}$$

where the projection P_i is defined by $P_i := |f_i\rangle\langle f_i| \otimes I_{B'}$. Since $\langle\tilde{f}_i|P_i|\tilde{f}_i\rangle = \langle\tilde{f}_i|f'_i\rangle\langle f'_i|\tilde{f}_i\rangle$,

$\mathrm{Tr}\,\rho(|e_i\rangle\langle e_i| \otimes |f_i\rangle\langle f_i|) = \langle\Phi|(|e_i\rangle\langle e_i| \otimes |f_i\rangle\langle f_i| \otimes I_{B'}|)|\Phi\rangle = \langle\Phi|(|e_i\rangle\langle e_i| \otimes |f'_i\rangle\langle f'_i|)|\Phi\rangle$.

Thus, the relations

$$T' := \sum_i p_i|e_i\rangle\langle e_i| \otimes |f'_i\rangle\langle f'_i| \le \sum_i p_i|e_i\rangle\langle e_i| \otimes |f_i\rangle\langle f_i| \otimes I_{B'} \le I_{ABB'}$$
$$\langle\Phi|T'|\Phi\rangle = \mathrm{Tr}\,\rho T = 1$$

hold. Moreover, T' satisfies the equality $\mathrm{Tr}\,T' = \sum_i p_i = \mathrm{Tr}\,T$. Thus, the inequality $t_{\widetilde{\rm sep}}(\rho) \ge t_{\widetilde{\rm sep}}(|\Phi\rangle)$ holds. Therefore, the relations $t_{\rm sep}(\rho) \ge t_{\widetilde{\rm sep}}(\rho) \ge t_{\widetilde{\rm sep}}(|\Phi\rangle) = (\mathrm{Tr}\sqrt{\rho_A})^2$ hold.

Similarly, we can show the inequality $t_{\rm sep}(\rho) \ge (\mathrm{Tr}\sqrt{\rho_B})^2$. Thus, we obtain (7) in the mixed states case. □

4. Local discrimination by one-way LOCC

In this section, we prove the following theorem concerning the local discrimination problem in terms of one-way LOCC in the direction $A \to B$:

Theorem 3 The inequality

$$t_{\to}(\rho) \ge \mathrm{rank}\,\rho_A \tag{22}$$

holds for a bipartite state ρ on $\mathcal{H}_A \otimes \mathcal{H}_B$. Any maximally correlated state ρ satisfies the equality. In other words, the following inequality concerning the minimum error probability $\beta_{\to}(\rho)$ holds:

$$\beta_{\to}(\rho) \ge \frac{1}{D}\mathrm{rank}\,\rho_A \tag{23}$$

In the above theorem, a maximally correlated state is defined as a state which can be decompose in the following form:

$$\rho = \sum_{1\le i,j\le d} \alpha_{ij}\,|u_i, v_i\rangle\,\langle u_j, v_j|\,, \tag{24}$$

where $\{|u_i\rangle\}_{i=1}^d$ and $\{|v_j\rangle\}_{j=1}^d$ are orthonormal bases of $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively [30]; apparently, a pure state is a maximally correlated state. Thus, $t_{\to}(|\Psi\rangle) = D\beta_{\to}(|\Psi\rangle)$ is

equal to the Schmidt rank of a state for a bipartite pure state $|\Psi\rangle$. In the case when ρ is a maximally correlated state satisfying Eq.(24), the optimal way to discriminate between ρ and the completely mixed state is the following: Suppose there are two parties called Alice and Bob. Both Alice and Bob measure their local states $\mathcal{H}_A$ and $\mathcal{H}_B$ in the bases $\{|u_i\rangle\}_{i=1}^d$ and $\{|v_j\rangle\}_{j=1}^d$, respectively, (of course, they only need to detect the support of the local states). Then, Alice informs her measurement result to Bob. Suppose Alice's result is $|u_k\rangle$ and Bob's result is $|v_l\rangle$. If k is equal to l, then, they judge that the given state is ρ. Otherwise, they judge that the given state is the completely mixed state.

By comparing Theorem 1 and Theorem 3, we can easily see that if a bipartite pure state $|\Psi\rangle$ is not a maximally entangled state nor a product state, then, the strict inequality $\beta_{\text{sep}}(|\Psi\rangle) < \beta_{\rightarrow}(|\Psi\rangle)$ holds. Thus, we can conclude that there is a gap between the one-way local distinguishability and the separable local distinguishability for a bipartite pure state at least in the present problem settings from these results.

Applying Theorem 3 for Eq.(6), we can extend Cohen's characterization [7] concerning the perfect discrimination of a given set of pure states in term of one-way LOCC to a set of mixed states:

Corollary 2 If a set of bipartite states $\{\rho_i\}_{i=1}^N$ on $\mathcal{H}_A \otimes \mathcal{H}_B$ is perfectly distinguishable by one-way LOCC, then,

$$\sum_{i=1}^N \text{rank}\rho_{iA} \leq D. \tag{25}$$

This bound of the size of locally distinguishable sets for one-way LOCC is much stronger than the known bound for separable operations [15].

As a preparation for our proof of Theorem 3, we see the fact that there are several equivalent representations of the definition of one-way LOCC POVM elements. We start from the following representation which we can see immediately from the definition; that is, in a bipartite system $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, if $T \in \mathcal{T}_{\rightarrow}$, there exist sets of positive operators $\{M_i\}_i$ and $\{N_j^i\}_j$ such that

$$T = \sum_{ij} M_i \otimes N_j^i, \tag{26}$$

$\sum_i M_i \leq I_B$, and $\sum_j N_j^i \leq I_A$, where $\{M_i\}_i$ is the POVM of the local measurement on $\mathcal{H}_A$ and $\{N_j^i\}_j$ is the POVM of the local measurement on $\mathcal{H}_B$ depending on the first measurement result i. Further, redefining N_j^i as $N_i = \sum_j N_j^i$, we derive the following equivalence relation:

$$\begin{aligned} & T \in \mathcal{T}_{\rightarrow} \\ \Longleftrightarrow\; & \exists \{M_i\}_i \text{ and } \{N_i\}_i \\ & s.t. \forall i, 0 \leq M_i, 0 \leq N_i \leq I_B, \sum_i M_i \leq I_A, \text{and } T = \sum_i M_i \otimes N_i. \end{aligned} \tag{27}$$

Using this characterization, we obtain the following lemma.

Lemma 2 A one-way LOCC POVM element $T = \sum_{ij} M_i \otimes N^i_j \in \mathcal{T}_{\rightarrow}$ satisfies $\mathrm{Tr}\,\rho T = 1$ if and only if $\mathrm{Tr}(\rho_A \sum M_i) = 1$ and $\mathrm{Tr}(\rho_{B,M_i} \sum_j N^i_j) = 1$ for all i, where $\rho_A \stackrel{\mathrm{def}}{=} \mathrm{Tr}_B\,\rho$ and $\rho_{B,M_i} \stackrel{\mathrm{def}}{=} \mathrm{Tr}_A\,\rho M_i \otimes I_B / \mathrm{Tr}\,\rho M_i \otimes I_B$.

Proof We can calculate $\mathrm{Tr}\,\rho T$ as follows:

$$\begin{aligned}\mathrm{Tr}\,\rho T = \sum_{ij} \mathrm{Tr}\,\rho M_i \otimes N^i_j =& \sum_{ij} \mathrm{Tr}\{(\mathrm{Tr}_A\,\rho(M_i \otimes I_B))N^i_j\} \\ =& \sum_i \mathrm{Tr}\,\rho_A M_i \cdot \mathrm{Tr}\,\rho_{B,M_i}(\sum_j N^j_i) = 1.\end{aligned}$$

Since $\sum_i \mathrm{Tr}\,\rho_A M_i \le 1$ and $\mathrm{Tr}\,\rho_{B,M_i}(\sum_j N^i_j) \le 1$ for all i, we derive $\sum_i \mathrm{Tr}\,\rho_A M_i = 1$ and $\mathrm{Tr}\,\rho_{B,M_i}(\sum_j N^i_j) = 1$. The opposite direction is trivial. □

Now, we prove Theorem 3 using the above lemma.

Proof (Theorem 3) In order to detect a state perfectly, we need to detect the reduced density operator of the local system A, ρ_A as well as that of the other local system B, ρ_{B,M_i}, perfectly in each step. Thus, we can assume that N_i is a projection on B without loss of generality. Hence, $\mathrm{Tr}\,T = \sum_i \mathrm{Tr}\,M_i \cdot \mathrm{Tr}\,N_i \ge \sum_i \mathrm{Tr}\,M_i$. Since we have to detect the reduced density operator of the local system A, ρ_A, we obtain $\mathrm{Tr}\sum_i M_i \rho_A = 1$, i.e., (22). When the state ρ is a maximally mixed state $\sum_{1\le i,j\le d} a_{i,j}|u_i, v_i\rangle\langle u_j, v_j|$, the reduced density ρ_A is $\sum_{i=1}^d a_{i,i}|u_i\rangle\langle u_i|$. Thus, $\mathrm{rank}\,\rho_A = d$. In this case, we can perfectly detect this state by the one-way LOCC test $\sum_{i=1}^d |u_i\rangle\langle u_i| \otimes |v_i\rangle\langle v_i|$. □

We should note the following fact: Although a maximally correlated state satisfies the equality of Eq.(22), the converse is not necessarily true. Even if v_i is not orthogonal, we can perfectly detect this state by the one-way LOCC test $\sum_{i=1}^d |u_i\rangle\langle u_i| \otimes |v_i\rangle\langle v_i|$. When the rank of the state $\sum_{1\le i,j\le d} a_{i,j}\,\langle v_j|v_i\rangle\,|u_i\rangle\langle u_j|$ is d, the rank of ρ_A is d. That is, this gives a counter example of the converse.

5. Local discrimination by two-way LOCC

So far, we have calculated the minimum error probability of the local discrimination problem for one-way LOCC $\beta_{\rightarrow}(|\Psi\rangle)$, and separable operations $\beta_{\mathrm{sep}}(|\Psi\rangle)$. In this section, we focus on discrimination protocols by two-way LOCC. Since the two-way LOCC is mathematically complicated, it is difficult to derive the minimum two-way LOCC discrimination protocol, and as a result, it is difficult to derive the exact value of $\beta_{\leftrightarrow}(|\Psi\rangle)$. However, in order to show the difference of the efficiency of one-way and two-way local discrimination protocols, (which is actually our main purpose of this paper,) it is enough to find the upper-bound of the two-way error probability $\beta_{\leftrightarrow}(|\Psi\rangle)$. Thus, we concentrate ourselves on driving an upper-bound of $\beta_{\leftrightarrow}$ by constructing a concrete two-way LOCC discrimination protocol. For simplicity, we only treat three-step LOCC discrimination protocols on a bipartite system, which are in the simplest class of genuine two-way LOCC protocols. As a result, we show that even three-step LOCC protocols

can discriminates a given state from the completely mixed state much better than by one-way (that is, two-step) LOCC protocols.

We can generally describe a three-step LOCC protocol to discriminate a pure state $|\Psi\rangle$ from $\rho_{\mathrm{mix}} = \frac{I}{d^2}$ on a bipartite system $\mathcal{H}_A \otimes \mathcal{H}_B$ without making any error to detect $|\Psi\rangle$ as follow: Suppose there are two parties called Alice and Bob. First, Alice performs a POVM $\{M_i\}_i$ on her system $\mathcal{H}_A$, and sends the measurement result i to Bob. Second, depending on i, Bob performs a POVM $\{N_j^i\}_j$ on his system $\mathcal{H}_B$, and sends the measurement result j to Alice. If the given state is ρ, by easy calculation, we can check that the Alice's state after this step is

$$\sigma_A^{ij} \stackrel{\mathrm{def}}{=} \frac{\sqrt{M_i}\sqrt{\rho_A}N_j^{iT}\sqrt{\rho_A}\sqrt{M_i}}{\mathrm{Tr}\left(M_i\sqrt{\rho_A}N_j^{iT}\sqrt{\rho_A}\right)}, \tag{28}$$

where $\rho_A \stackrel{\mathrm{def}}{=} \mathrm{Tr}_B |\Psi\rangle\langle\Psi|$, and the transposition is taken in the Schmidt basis of $|\Psi\rangle$. Thus, in order not to make an error to detect the above state, finally, Alice should make a measurement in $\{\{\sigma_A^{ij} > 0\}, I_A - \{\sigma_A^{ij} > 0\}\}$, where $\{\sigma_A^{ij} > 0\}$ is a projection operator onto the support of σ_A^{ij} (the subspace spanned by eigenvectors corresponding to non-zero eigenvalue of σ_A^{ij}), I_A is an identity operator in $\mathcal{H}_A$. Then, if she detects $\{\sigma_A^{ij} > 0\}$, she judges that the first state was $|\Psi\rangle$, and if she detects $I_A - \{\sigma_A^{ij} > 0\}$, she judges that the first state was ρ_{mix}. Suppose $\{T, I_{AB} - T\}$ is the POVM corresponds to the above local discrimination protocol, where T corresponds to $|\Psi\rangle$, and $I - T$ corresponds to ρ_{mix} Then, we can check that the whole POVM $\{T, I - T\}$ can be written down as follows,

$$T = \sum_i \sum_j \left(\sqrt{M_i}\{\sigma_A^{ij} > 0\}\sqrt{M_i}\right) \otimes N_j^i, \tag{29}$$

where σ_A^{ij} is defined by Eq.(28), and all M_i and N_j^i are positive operators satisfying $\sum_i M_i = I_A$ and $\sum_j N_j^i = I_B$. We can also check that T defined by Eq.(29) satisfies $\langle\Psi| T |\Psi\rangle = 1$ as follows:

$$\begin{aligned}
\langle\Psi| T |\Psi\rangle \quad &= \sum_{ij} \langle\Psi| \left(\sqrt{M_i}\{\sigma_A^{ij} > 0\}\sqrt{M_i}\right) \otimes N_j^i |\Psi\rangle \\
&= d\sum_{ij} \left(\langle\Phi^+| \sqrt{\rho_A} \otimes I_B\right) \left(\sqrt{M_i}\{\sigma_A^{ij} > 0\}\sqrt{M_i}\right) \otimes N_j^i \left(\sqrt{\rho_A} \otimes I_B |\Phi^+\rangle\right) \\
&= d\sum_{ij} \langle\Phi^+| \left(\sqrt{N_j^i}^T \sqrt{\rho_A}\sqrt{M_i}\{\sigma_A^{ij} > 0\}\sqrt{M_i}\sqrt{\rho_A}\sqrt{N_j^i}^T\right) \otimes I_B |\Phi^+\rangle \\
&= \sum_{ij} \mathrm{Tr} \sqrt{N_j^i}^T \sqrt{\rho_A}\sqrt{M_i}\{\sigma_A^{ij} > 0\}\sqrt{M_i}\sqrt{\rho_A}\sqrt{N_j^i}^T \\
&= \sum_{ij} \mathrm{Tr} \sqrt{M_i}\sqrt{\rho_A}N_j^{iT}\sqrt{\rho_A}\sqrt{M_i}\{\sigma_A^{ij} > 0\} \\
&= \sum_{ij} \left(\mathrm{Tr}\, M_i\sqrt{\rho_A}N_j^{iT}\sqrt{\rho_A}\right) \cdot \left(\mathrm{Tr}\, \sigma_A^{ij}\{\sigma_A^{ij} > 0\}\right) \\
&= \sum_{ij} \mathrm{Tr} \left(N_j^i \left(\sqrt{\rho_A}M_i\sqrt{\rho_A}\right)^T\right)
\end{aligned}$$

$$= \sum_{ij} \mathrm{Tr}\, \rho_A M_i$$
$$= 1,$$

where $|\Phi^+\rangle$ is the maximally entangled state sharing the Schmidt basis with $|\Psi\rangle$, and the transposition T is always taken in the Schmidt basis of $|\Psi\rangle$. In the second line of the above equalities, we used the equality $|\Psi\rangle = \sqrt{\rho_A} \otimes I_B |\Phi^+\rangle$. In the third line, we used the equalities $I_A \otimes X |\Phi^+\rangle = X^T \otimes I_B |\Phi^+\rangle$, which is valid for an arbitrary operator X. In the sixth line, we used Eq.(28).

The above three-step LOCC protocol is enough general. However, it is too complicated to optimize $\mathrm{Tr}\, T$ over all choices of POVM $\{M_i\}_i$ and $\{N^i_j\}_j$. In this section, our aim is only to find a good (not necessary optimal) two-way LOCC protocol by which we can discriminate a state from the completely mixed state better than by any one-way LOCC protocols. Thus, to make a problem simpler, we make the following assumptions on Alice's POVM $\{M_i\}_i$ and Bob's POVM $\{N^i_j\}_j$: First, without losing any generality, we can write $|\Psi\rangle$ as $|\Psi\rangle = \sum_{k=1}^{d} \sqrt{\lambda_k}\, |k\rangle \otimes |k\rangle$ with $\lambda_i > 0$ and $\lambda_i \geq \lambda_{i+1}$, where $\{|k\rangle\}_{k=1}^{d}$ is an arbitrary fixing computational basis of $\mathcal{H}_A$ and $\mathcal{H}_B$, and $d = \dim \mathcal{H}_A = \dim \mathcal{H}_B$; since our definition of $t_{\leftrightarrow}(|\Psi\rangle)$ (Eq.(4)) does not depend on the dimension of the whole system, we can always choose the whole system in order that the Schmidt rank of $|\Psi\rangle$ is d. Second, we assume that the number of POVM element M_i is d, and M_i is diagonalizable in the computational basis $\{|k\rangle\}_{k=1}^{d}$ as,

$$M_i = \sum_{k=1}^{i} \delta_{ki} |k\rangle \langle k| , \tag{30}$$

where $\mathrm{rank} M_i = i$, and the coefficients $\delta_{ki} \geq 0$ satisfy $\sum_{i=k}^{d} \delta_{ki} = 1$ for all k. Moreover, we assume that $\{N^i_j\}_j$ is a von Neumann measurement corresponding to a mutually unbiased basis [11, 31] $\{|\xi^i_j\rangle\}_{j=1}^{\mathrm{rank} M_i}$ of an orthonormal set of eigen vectors of $\omega_B \stackrel{\mathrm{def}}{=} \frac{\sqrt{\rho_B} M_i^T \sqrt{\rho_B}}{\mathrm{Tr}\, \sqrt{\rho_B} M_i^T \sqrt{\rho_B}}$ corresponding to non-zero eigenvalues; that is, $\{\xi^i_j\}_{j=1}^{\mathrm{rank} M_i}$ only spans $\mathrm{Ran} \frac{\sqrt{\rho_B} M_i^T \sqrt{\rho_B}}{\mathrm{Tr}\, \sqrt{\rho_B} M_i^T \sqrt{\rho_B}}$. In other words, an orthonormal set of states $\{|\xi^i_j\rangle\}_{j=1}^{\mathrm{rank} M_i}$ satisfies

$$\langle \xi^i_j | \frac{\sqrt{\rho_B} M_i^T \sqrt{\rho_B}}{\mathrm{Tr}\, \sqrt{\rho_B} M_i^T \sqrt{\rho_B}} |\xi^i_j\rangle = \frac{1}{\mathrm{rank} M_i}. \tag{31}$$

Note that ω_B is the Bob's state after the Alice's first measurement in the case where the given state is $|\Psi\rangle$, and thus, Bob only needs to detect the subspace $\mathrm{Ran}\omega_B$ in this case. That is, Bob's POVM consists of $\{|\xi^i_j\rangle \langle \xi^i_j|\}_{j=1}^{\mathrm{rank} M_i}$ and $I_B - \sum_{j=1}^{\mathrm{rank} M_i} |\xi^i_j\rangle \langle \xi^i_j|$; if Bob derives the measurement result corresponding to $I_B - \sum_{j=1}^{\mathrm{rank} M_i} |\xi^i_j\rangle \langle \xi^i_j|$, then, he judges that the given state is ρ_{mix}. We also should note that due to this Bob's mutually unbiased measurement, our three-step protocol can not be reduced to a two-step one-way LOCC protocol. If all Bob's POVMs are commutative with the eigen basis of ω_B, the whole protocol can be reduced to a one-way LOCC protocol; however, ω_B never commutes the projection onto the mutually unbiased basis of the eigen basis of ω_B. Finally, under the above assumptions, we can write down the trace of the whole POVM

element $\mathrm{Tr}\, T$ as follows,

$$
\begin{aligned}
\mathrm{Tr}\, T =\ & \mathrm{Tr} \left(\sum_{i=1}^{d} \sum_{j=1}^{\mathrm{rank} M_i} \left(\sqrt{M_i} \{ \sigma_A^{ij} > 0 \} \sqrt{M_i} \right) \otimes N_j^i \right) \\
=\ & \sum_{i=1}^{d} \sum_{j=1}^{\mathrm{rank} M_i} \mathrm{Tr} \left(\left(\sqrt{M_i} \left(\frac{\sqrt{M_i}\sqrt{\rho_A}(|\xi_j^i\rangle \langle \xi_j^i|)^T \sqrt{\rho_A}\sqrt{M_i}}{\langle \xi_j^i | \sqrt{\rho_A} M_i^T \sqrt{\rho_A} | \xi_j^i \rangle} \right) \sqrt{M_i} \right) \otimes |\xi_j^i\rangle \langle \xi_j^i| \right) \\
=\ & \sum_{i=1}^{d} \sum_{j=1}^{\mathrm{rank} M_i} \frac{\langle \xi_j^i | \sqrt{\rho_A} (M_i^T)^2 \sqrt{\rho_A} | \xi_j^i \rangle}{\langle \xi_j^i | \sqrt{\rho_A} M_i^T \sqrt{\rho_A} | \xi_j^i \rangle} \\
=\ & \sum_{i=1}^{d} \mathrm{rank} M_i \frac{\mathrm{Tr} \sqrt{\rho_A} (M_i^T)^2 \sqrt{\rho_A}}{\mathrm{Tr} \sqrt{\rho_A} M_i^T \sqrt{\rho_A}} \\
=\ & \sum_{i=1}^{d} i \cdot \frac{\sum_{k=1}^{i} \lambda_k \delta_{ki}^2}{\sum_{k=1}^{i} \lambda_i \delta_{ki}}.
\end{aligned}
$$

In the second line of the above equalities, we used Eq.(28) (the definition of σ_A^{ij}) and the equality $\left\{ \frac{\sqrt{M_i}\sqrt{\rho_A}(|\xi_j^i\rangle \langle \xi_j^i|)^T \sqrt{\rho_A}\sqrt{M_i}}{\langle \xi_j^i | \sqrt{\rho_A} M_i^T \sqrt{\rho_A} | \xi_j^i \rangle} > 0 \right\} = \frac{\sqrt{M_i}\sqrt{\rho_A}(|\xi_j^i\rangle \langle \xi_j^i|)^T \sqrt{\rho_A}\sqrt{M_i}}{\langle \xi_j^i | \sqrt{\rho_A} M_i^T \sqrt{\rho_A} | \xi_j^i \rangle}$. In the fourth line of the above equalities, we used the relation $\rho_A = \rho_B$ and the condition of mutually unbiased basis Eq.(31). Therefore, our problem is reduced to the optimization of $\sum_{i=1}^{d} i \cdot \frac{\sum_{k=1}^{i} \lambda_k \delta_{ki}^2}{\sum_{k=1}^{i} \lambda_i \delta_{ki}}$ over $\{\delta_{ki}\}_{ki}$ subjected to the constraints $\delta_{ki} \geq 0$ and $\sum_{i=k}^{d} \delta_{ki} = 1$. In other words, we can summarized the above discussion in the form of the following lemma.

Lemma 3 For a bipartite pure state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, $\beta_{\leftrightarrow}(|\Psi\rangle)$ satisfies the following inequality,

$$
\begin{aligned}
& \beta_{\leftrightarrow}(|\Psi\rangle) \leq \beta_{\rightleftarrows}(|\Psi\rangle) \\
& \quad \stackrel{\mathrm{def}}{=} \frac{1}{d^2} \min_{\{\delta_{ki}\}_{1\leq k \leq i \leq d}} \left\{ \sum_{i=1}^{d} i \cdot \frac{\sum_{k=1}^{i} \lambda_k \delta_{ki}^2}{\sum_{k=1}^{i} \lambda_k \delta_{ki}} \,\middle|\, \forall k, \forall i, \delta_{ki} \geq 0, \text{ and } \forall k, \sum_{i=k}^{d} \delta_{ki} = 1 \right\}, \qquad (32)
\end{aligned}
$$

where $\{\lambda_k\}_{k=1}^d$ is the Schmidt coefficients of $|\Psi\rangle$, and satisfies $\lambda_k \geq \lambda_{k+1}$ for all k, and the indices (k,i) moves among all of the triangle region $1 \leq k \leq i \leq d$.

Then, the above inequality can write as $\beta_{\leftrightarrow}(|\Psi\rangle) \leq \beta_{\rightleftarrows}(|\Psi\rangle)$. Together with the results of the past sections, we derive the following inequalities related to the minimum type 2 error probability for a bipartite pure state:

$$
\begin{aligned}
& \beta_g(|\Psi\rangle) = \frac{1}{d^2} \leq \beta_{\mathrm{sep}}(|\Psi\rangle) = \frac{1}{d^2} \Big(\sum_i \sqrt{\lambda_i} \Big)^2 \\
& \qquad \leq \beta_{\leftrightarrow}(|\Psi\rangle) \leq \beta_{\rightleftarrows}(|\Psi\rangle) \leq \frac{1}{d^2} \mathrm{rank}\, \mathrm{Tr}_B (|\Psi\rangle \langle \Psi|) = \beta_{\rightarrow}(|\Psi\rangle). \qquad (33)
\end{aligned}
$$

For a two-qubit system, we can analytically calculate the exact value of the upper-bound $\beta_{\rightleftarrows}(|\Psi\rangle)$, and can derive the following lemma.

Lemma 4 In a two-qubit system,

$$\beta_{\rightleftarrows}(|\Psi\rangle) = \frac{1}{2} - \frac{(1-\sqrt{2\lambda})^2}{4\,(1-\lambda)}, \tag{34}$$

where $\{1-\lambda, \lambda\}$ is the Schmidt coefficient of $|\Psi\rangle$ satisfying $1 \leq \lambda \leq \frac{1}{2}$.

Proof Without losing generality, we can write a bipartite state as $|\Psi\rangle = \sqrt{1-\lambda}\,|00\rangle + \sqrt{\lambda}\,|11\rangle$. Then, by a straightforward calculation, we derive

$$\beta_{\rightleftarrows}(|\Psi\rangle) = \frac{1}{4}\min_{0\leq\delta\leq1}\left\{2 - \frac{\delta\,\{(1-2\lambda)+\delta(1-\lambda)\}}{1-\delta(1-\lambda)}\right\}, \tag{35}$$

where we substitute $\lambda_1 = 1-\lambda$, $\lambda_2 = \lambda$, $\delta_{11} = \delta$, $\delta_{12} = 1-\delta$, and $\delta_{22} = 1$ into Eq.(32). Suppose $t'_{\rightleftarrows}(\lambda,\delta) \stackrel{\text{def}}{=} 2 - \frac{\delta\{(1-2\lambda)+\delta(1-\lambda)\}}{1-\delta(1-\lambda)}$. Then, we can calculate the derivative of $t'_{\rightleftarrows}(\lambda,\delta)$ as

$$\frac{\partial t'_{\rightleftarrows}(\lambda,\delta)}{\partial\delta} = -\{(1-\lambda)\delta - (1-\sqrt{2\lambda})\}\{(1-\lambda)\delta - (1+\sqrt{2\lambda})\},$$

for fixed $0 \leq \lambda \leq \frac{1}{2}$. Thus, under the condition $0 \leq \delta \leq 1$, $t'_{\rightleftarrows}(\lambda,\delta)$ attains its minimum when $\delta = \frac{1-\sqrt{2\lambda}}{1-\lambda}$. Therefore, we derive

$$\beta_{\rightleftarrows}(|\Psi\rangle) = \frac{1}{4}\min_{0\leq\delta\leq1} t'_{\rightleftarrows}(\lambda,\delta) = \frac{1}{2} - \frac{(1-\sqrt{2\lambda})^2}{4\,(1-\lambda)}.$$

$\square$

Therefore, for a two-qubit state $|\Psi_\lambda\rangle = \sqrt{1-\lambda}\,|00\rangle + \sqrt{\lambda}\,|11\rangle$, the inequality (33) can be reduced as follows,

$$\begin{aligned}\beta_g(|\Psi\rangle) = \frac{1}{4} \leq \beta_{\text{sep}}(|\Psi\rangle) = \frac{1}{4} + \frac{1}{2}\sqrt{\lambda(1-\lambda)} \leq \beta_{\leftrightarrow}(|\Psi\rangle) &\leq \frac{1}{2} - \frac{(1-\sqrt{2\lambda})^2}{4\,(1-\lambda)} \\ &\leq \frac{1}{2} = \beta_{\rightarrow}(|\Psi\rangle),\end{aligned}$$

where the equality of the last inequality holds, if and only if the state is a product state or a maximally entangled state. We present the graph of these bounds in **Figure.1**. From this figure, we can see that there is a big gap between $\beta_{\rightarrow}(|\Psi\rangle)$ and $\beta_{\leftrightarrow}(|\Psi\rangle)$ and the difference between $\beta_{\rightarrow}(|\Psi\rangle)$ and $\beta_{\text{sep}}(|\Psi\rangle)$ is (if the difference exists) relatively small. Thus, for any non-maximally entangled pure states, there is a gap between the one-way and two-way local distinguishability at least for two-qubit systems in terms of $\beta_{\rightarrow(\leftrightarrow)}(|\Psi\rangle)$.

In a system with a dimension of local systems $d \geq 3$, the optimization in the definition of $\beta_{\rightleftarrows}(|\Psi\rangle)$ (Eq.(32)) is too complicated to be solved by an analytical calculation, anymore ‡. Thus, we numerically calculate the right hand side of Eq.(32) for a $\mathbb{C}^3 \otimes \mathbb{C}^3$ (two-qutrit) system and a $\mathbb{C}^4 \otimes \mathbb{C}^4$ system. For a $\mathbb{C}^3 \otimes \mathbb{C}^3$ system, we calculate Eq.(32) for three different one-parameter families of pure states:

‡ In a strict sense , we can show that there exists an analytical solution for the optimization problem in Eq.(32) by means of Lagrange multiplier. However, even for a 3×3 dimensional system, the general solution is too complicated and too ugly to write here.

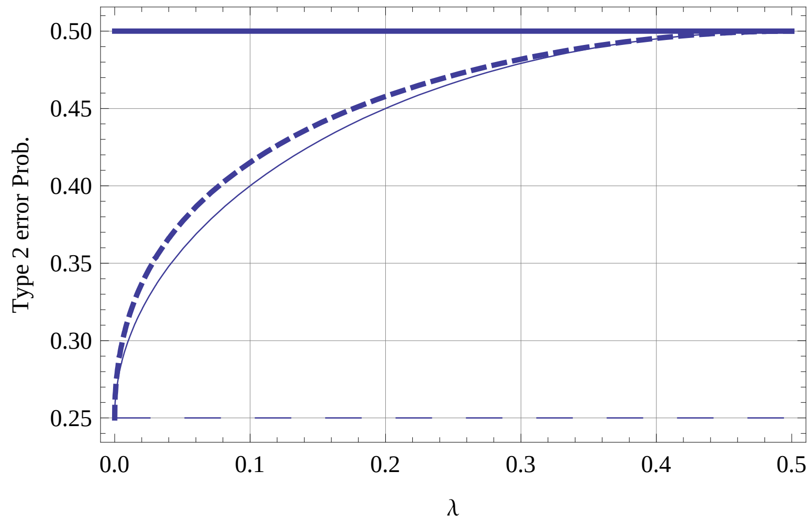

Figure 1. The bound as a function of λ (the Schmidt coefficient of $|\Psi\rangle$). The thin line: $\beta_{\rm sep}(|\Psi\rangle)$, the broken line: $\beta_{\rightleftarrows}(|\Psi\rangle)$ (an upper bound of $\beta_{\leftrightarrow}(|\Psi\rangle)$), the thick line: $\beta_{\rightarrow}(|\Psi\rangle)$, the thin broken line: $\beta_g(|\Psi\rangle)$.

(i) $|\Psi_\lambda\rangle = \sqrt{1-2\lambda}\,|11\rangle + \sqrt{\lambda}\,|22\rangle + \sqrt{\lambda}\,|33\rangle$, $(0 \le \lambda \le \frac{1}{3})$: In this case, $\beta_g(|\Psi_\lambda\rangle) = \frac{1}{9}$, $\beta_{\rm sep}(|\Psi_\lambda\rangle) = \frac{1}{9}(\sqrt{1-2\lambda} + 2\sqrt{\lambda})^2$ and $\beta_{\rightarrow}(|\Psi_\lambda\rangle) = \frac{1}{3}$. We give the results of numerical calculation of $\beta_{\widetilde{\rm sep}}(|\Psi_\lambda\rangle)$ in **Figure.2**.

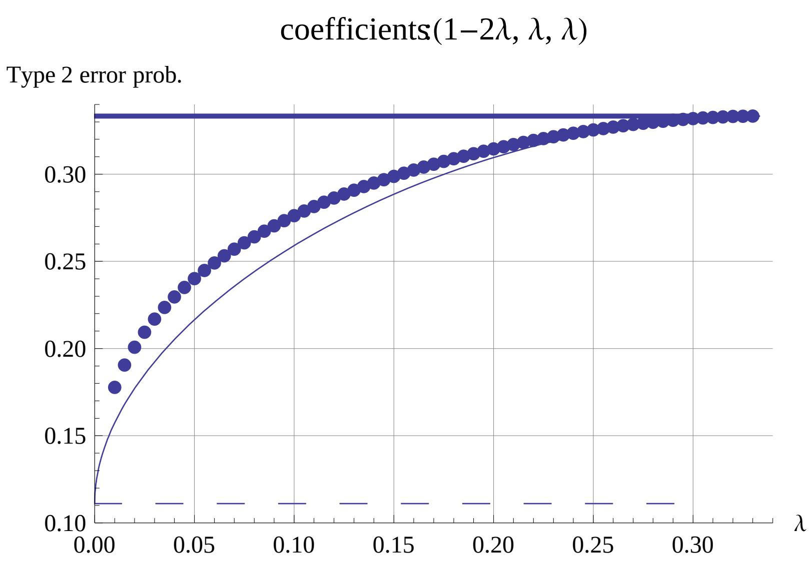

Figure 2. The bound as a function of λ for a family of states $|\Psi_\lambda\rangle = \sqrt{1-2\lambda}\,|11\rangle + \sqrt{\lambda}\,|22\rangle + \sqrt{\lambda}\,|33\rangle$. The thick broken line: results of a numerical calculation of $\beta_{\rightleftarrows}(|\Psi_\lambda\rangle)$ (a lower bound of $\beta_{\leftrightarrow}(|\Psi_\lambda\rangle)$, the thin line: $\beta_{\rm sep}(|\Psi_\lambda\rangle)$, the thick line: $\beta_{\rightarrow}(|\Psi_\lambda\rangle)$: the thin broken line: $\beta_g(|\Psi_\lambda\rangle)$.

(ii) $|\Psi_\lambda\rangle = \sqrt{1-3\lambda}\,|11\rangle + \sqrt{2\lambda}\,|22\rangle + \sqrt{\lambda}\,|33\rangle$, $(0 \le \lambda \le \frac{1}{5})$: In this case, $\beta_g(|\Psi_\lambda\rangle) = \frac{1}{9}$, $\beta_{\rm sep}(|\Psi_\lambda\rangle) = \frac{1}{9}(\sqrt{1-3\lambda} + (1+\sqrt{2})\sqrt{\lambda})^2$ and $\beta_{\rightarrow}(|\Psi_\lambda\rangle) = \frac{1}{3}$. We give the results of

a numerical calculation of $\beta_{\widetilde{\rm sep}}(|\Psi_\lambda\rangle)$ in **Figure.3**

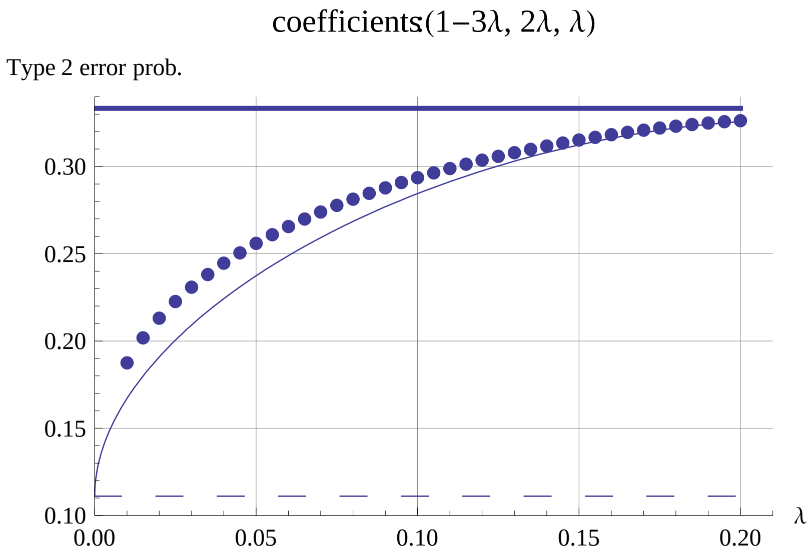

Figure 3. The bound as a function of λ for a family of states $|\Psi_\lambda\rangle = \sqrt{1-3\lambda}\,|11\rangle + \sqrt{2\lambda}\,|22\rangle + \sqrt{\lambda}\,|33\rangle$. The thick broken line: results of a numerical calculation of $\beta_{\rightleftarrows}(|\Psi_\lambda\rangle)$ (a lower bound of $\beta_{\leftrightarrow}(|\Psi_\lambda\rangle)$, the thin line: $\beta_{\rm sep}(|\Psi_\lambda\rangle)$, the thick line: $\beta_{\rightarrow}(|\Psi_\lambda\rangle)$, the thin broken line: $\beta_g(|\Psi_\lambda\rangle)$.

(iii) $|\Psi_\lambda\rangle = \sqrt{1-4\lambda}\,|11\rangle + \sqrt{3\lambda}\,|22\rangle + \sqrt{\lambda}\,|33\rangle$, $(0 \le \lambda \le \frac{1}{7})$: In this case, $\beta_g(|\Psi_\lambda\rangle) = \frac{1}{9}$, $\beta_{\rm sep}(|\Psi_\lambda\rangle) = \frac{1}{9}(\sqrt{1-4\lambda} + (1+\sqrt{3})\sqrt{\lambda})^2$ and $\beta_{\rightarrow}(|\Psi_\lambda\rangle) = \frac{1}{3}$. We give the results of a numerical calculation of $\beta_{\widetilde{\rm sep}}(|\Psi_\lambda\rangle)$ in **Figure.4**

From **Figures 2, 3, and 4**, we can confirm that the shapes of the graphs of $\beta_{\rm sep}(|\Psi_\lambda\rangle)$, and $\beta_{\rightleftarrows}(|\Psi_\lambda\rangle)$ hardly depend of the choice of a one-parameter family $|\Psi_\lambda\rangle$ in $\mathbb{C}^3 \otimes \mathbb{C}^3$. For a $\mathbb{C}^4 \otimes \mathbb{C}^4$ system, we calculate Eq.(32) for two different one-parameter families of pure states:

(i) $|\Psi_\lambda\rangle = \sqrt{1-3\lambda}\,|11\rangle + \sqrt{\lambda}(|22\rangle + |33\rangle + |44\rangle)$, $(0 \le \lambda \le \frac{1}{4})$: In this case, $\beta_g(|\Psi_\lambda\rangle) = \frac{1}{16}$, $\beta_{\rm sep}(|\Psi_\lambda\rangle) = \frac{1}{16}(\sqrt{1-3\lambda} + 3\sqrt{\lambda})^2$ and $\beta_{\rightarrow}(|\Psi_\lambda\rangle) = \frac{1}{4}$. We give the results of numerical calculation of $\beta_{\widetilde{\rm sep}}(|\Psi_\lambda\rangle)$ in **Figure.5**.

(ii) $|\Psi_\lambda\rangle = \sqrt{1-\frac{9}{2}\lambda}\,|11\rangle + \sqrt{2\lambda}\,|22\rangle + \sqrt{\frac{3}{2}\lambda}\,|33\rangle + \sqrt{\lambda}\,|44\rangle$, $(0 \le \lambda \le \frac{2}{13})$: In this case, $\beta_g(|\Psi_\lambda\rangle) = \frac{1}{16}$, $\beta_{\rm sep}(|\Psi_\lambda\rangle) = \frac{1}{16}\left(\sqrt{1-\frac{9}{2}\lambda} + \left(1+\sqrt{\frac{3}{2}}+\sqrt{2}\right)\sqrt{\lambda}\right)^2$ and $\beta_{\rightarrow}(|\Psi_\lambda\rangle) = \frac{1}{4}$. We give the results of numerical calculation of $\beta_{\widetilde{\rm sep}}(|\Psi_\lambda\rangle)$ in **Figure.6**

From **Figures. 5, and 6**, we can confirm that the shapes of the graphs of $\beta_{\rm sep}(|\Psi_\lambda\rangle)$, and $\beta_{\rightleftarrows}(|\Psi_\lambda\rangle)$ hardly depend of the choice of a one-parameter family $|\Psi_\lambda\rangle$ in $\mathbb{C}^4 \otimes \mathbb{C}^4$ as well as in $\mathbb{C}^3 \otimes \mathbb{C}^3$. Note that, for the all above families of states, we choose a parameter λ so that $|\Psi_\lambda\rangle$ can be converted to $|\Psi_{\lambda'}\rangle$ by LOCC for all $\lambda \ge \lambda'$, and $|\Psi_0\rangle$ is a product state; that is, in a naive sense, the degree of entanglement increases monotonically when

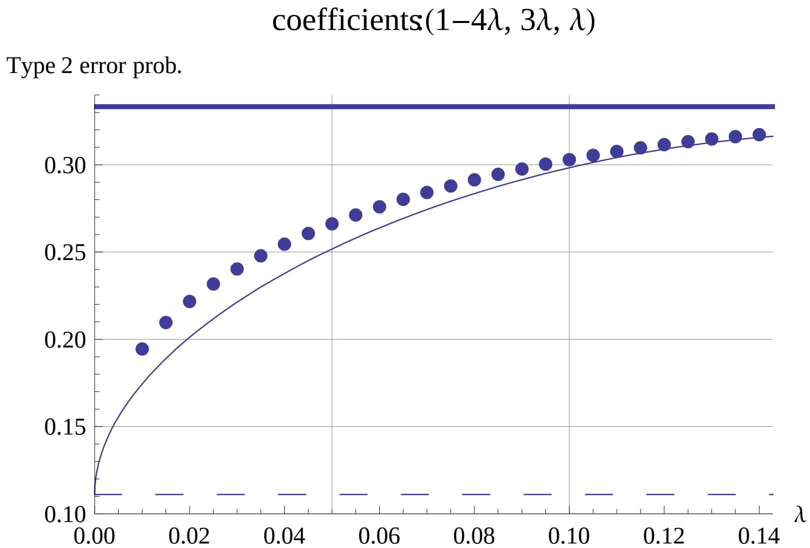

Figure 4. The bound as a function of λ for a family of states $|\Psi_\lambda\rangle = \sqrt{1-4\lambda}\,|11\rangle + \sqrt{3\lambda}\,|22\rangle + \sqrt{\lambda}\,|33\rangle$. The thick broken line: results of a numerical calculation of $\beta_{\rightleftarrows}(|\Psi_\lambda\rangle)$ (a lower bound of $\beta_{\leftrightarrow}(|\Psi_\lambda\rangle)$, the thin line: $\beta_{\mathrm{sep}}(|\Psi_\lambda\rangle)$, the thick line: $\beta_{\rightarrow}(|\Psi_\lambda\rangle)$, the thin broken line: $\beta_g(|\Psi_\lambda\rangle)$.

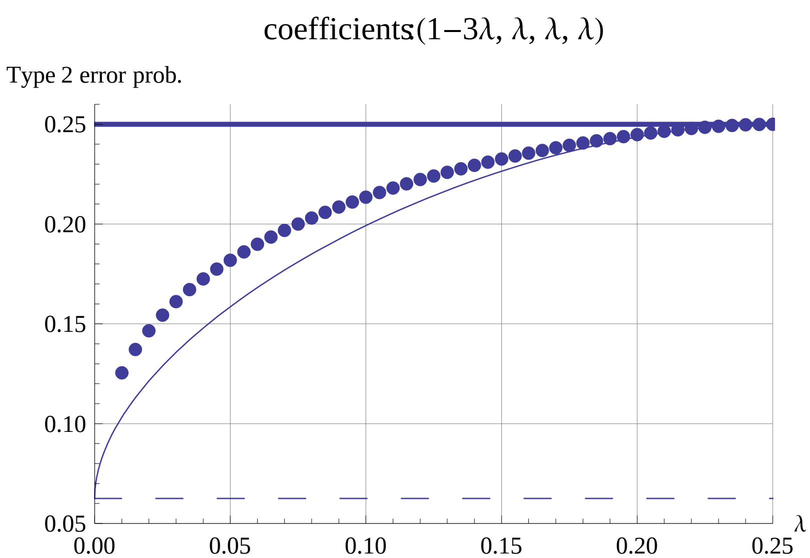

Figure 5. The bound as a function of λ for a family of states $|\Psi_\lambda\rangle = \sqrt{1-3\lambda}\,|11\rangle + \sqrt{\lambda}(|22\rangle + |33\rangle + |44\rangle)$. The thick broken line: results of a numerical calculation of $\beta_{\rightleftarrows}(|\Psi_\lambda\rangle)$ (a lower bound of $\beta_{\leftrightarrow}(|\Psi_\lambda\rangle)$, the thin line: $\beta_{\mathrm{sep}}(|\Psi_\lambda\rangle)$, the thick line: $\beta_{\rightarrow}(|\Psi_\lambda\rangle)$, the thin broken line: $\beta_g(|\Psi_\lambda\rangle)$.

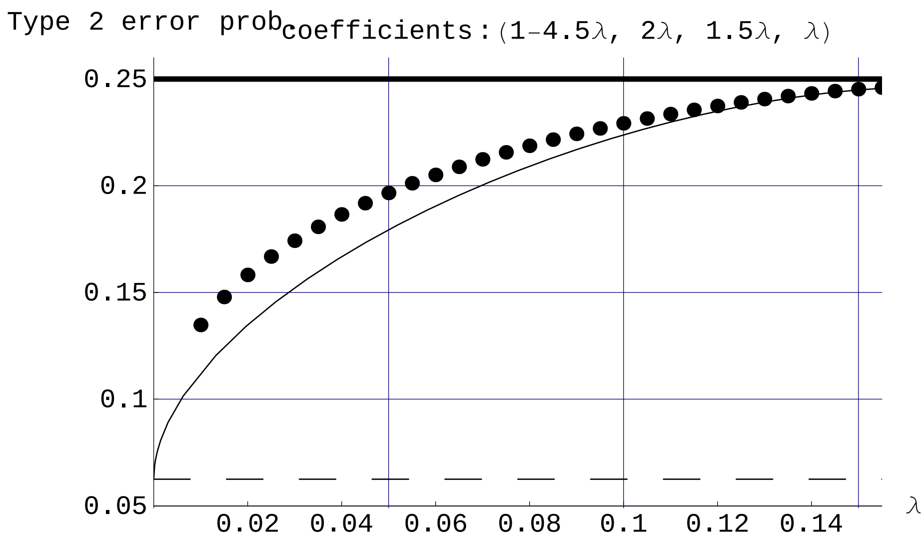

Figure 6. The bound as a function of λ for a family of states $|\Psi_\lambda\rangle = |\Psi_\lambda\rangle = \sqrt{1-\frac{9}{2}\lambda}\,|11\rangle + \sqrt{2\lambda}\,|22\rangle + \sqrt{\frac{3}{2}\lambda}\,|33\rangle + \sqrt{\lambda}\,|44\rangle$. The thick broken line: results of a numerical calculation of $\beta_{\rightleftarrows}(|\Psi_\lambda\rangle)$ (a lower bound of $\beta_{\leftrightarrow}(|\Psi_\lambda\rangle)$), the thin line: $\beta_{\mathrm{sep}}(|\Psi_\lambda\rangle)$, the thick line: $\beta_{\rightarrow}(|\Psi_\lambda\rangle)$, the thin broken line: $\beta_g(|\Psi_\lambda\rangle)$.

λ increases. From **Figures 2,3,4,5,6**, as well as for a two-qubit system (**Figure.1**), we can see that there is always a big gap between $\beta_{\rightarrow}(|\Psi\rangle)$ and $\beta_{\leftrightarrow}(|\Psi\rangle)$ and the difference between $\beta_{\rightarrow}(|\Psi\rangle)$ and $\beta_{\mathrm{sep}}(|\Psi\rangle)$ is (if the difference exists) relatively small for $\mathbb{C}^3 \otimes \mathbb{C}^3$ and $\mathbb{C}^4 \otimes \mathbb{C}^4$ systems. Moreover, since the shape of graph corresponding to $\beta_{\rightleftarrows}(|\Psi\rangle)$ seems not to change depending on a dimension of a system, we may guess that, for any non-maximally entangled pure states (even in a high dimensional system), there is a gap between the one-way and two-way local distinguishability in terms of $\beta_{\rightarrow(\leftrightarrow)}(|\Psi\rangle)$. That is, the two-way classical communication remarkably improves the local distinguishability compared to the local discrimination by the one-way classical communication at least for bipartite pure states.

6. Conclusion

In this paper, in order to clarify the difference of the two-way LOCC and the one-way LOCC on local discrimination problems, we concentrated ourselves on the local discrimination of a given bipartite state from the completely mixed state ρ_{mix} under the condition where the given state should be detected perfectly while the previous researches [11, 12] treated the same problem between two bipartite pure states. We defined $\beta_{\rightarrow}(\rho)$, $\beta_{\leftrightarrow}(\rho)$, and $\beta_{\mathrm{sep}}(\rho)$ as the minimum error probability to detect the completely mixed state by the one-way LOCC, the two-way LOCC, and the separable operation, respectively, under the condition that a given state ρ is detected perfectly. First, in Section 3, for separable operations, we showed that the minimum error

probability $t_{\rm sep}(\rho)$ coincides with an entanglement measure called the global robustness of entanglement for a bipartite pure state except an unimportant constant term. Then, in Section 4, for one-way LOCC, we showed that the minimum error probability $\beta_{\rightarrow}(\rho)$ coincides with the Schmidt rank for a bipartite pure state except an unimportant constant term. Finally, in Section 5, by constructing a concrete three-step two-way LOCC discrimination protocol, we derived an upper bound for the minimum error probability $\beta_{\leftrightarrow}(\rho)$ for a bipartite pure state. By calculating this upper bound analytically and also numerically, we showed that $\beta_{\leftrightarrow}(\rho)$ is strictly smaller than $\beta_{\rightarrow}(\rho)$, and moreover, $\beta_{\leftrightarrow}(\rho)$ and $\beta_{\rm sep}(\rho)$ give almost the same value for a lower dimensional bipartite pure state; this results can be seen in Figures 2,3,4,5,6. As a result, although there is no difference between the one-way LOCC and the two-way LOCC concerning local discrimination between two bipartite pure states [11, 12], we conclude that the two-way classical communication remarkably improves the local distinguishability in comparison with the one-way classical communication for a low-dimensional pure state at least in the present problem setting. Due to our quantitative comparison, from the continuity of the second kinds of error probabilities, a similar result should holds when the second state $\tilde{\rho}$ belongs to the neighborhood of the completely mixed state. Further, we are preparing a forthcoming manuscript concerning this kind of problem in the case of multi-partite case in the near future [32].

Acknowledgments

MO would like to thank Prof. Mio Murao, Dr. Damian Markham, and Dr. Shashank Virmani for helpful comments and discussions. MO is grateful to Professor Hiroshi Imai for his support and encouragement The authors also appreciate Reviewr's comments.

This work was supported by Special Coordination Funds for Promoting Science and Technology and a MEXT Grant-in-Aid for Scientific Research on Priority, Deepening and Expansion of Statistical Mechanical Informatics (DEX-SMI), No. 18079014.

References

[1] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher *Phys. Rev. A* **53**, 2046-2052 (1996); M. A. Nielsen, *Phys. Rev. Lett.* **83** 436 (1999); G. Vidal, *Phys. Rev. Lett.* **83**, 1046 (1999).

[2] H. K. Lo, and S. Popescu, *Phys. Rev. A*, 63, 022301 (2001)

[3] M. Hayashi, *Quantum Information: An Introduction*, Springer-Verlag, (2006)

[4] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, *Phys. Rev. A* **54**, 3824 (1996)

[5] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters *Phys. Rev. A* **59**, 1070-1091 (1999); S. Ghosh, G. Kar, A. Roy, and D. Sarkar, *Phys. Rev. A* **70**, 022304 (2004) ; H. Fan, *Phys. Rev. Lett.* **92**, 177905 (2004); M. Nathanson, *J. Math. Phys.* 46, 062103 (2005)

[6] B. Groisman and L. Vaidman, *J. Phys. A*, **34**, 6881 (2001).

[7] S. M. Cohen, *Phy. Rev. A*, **75**, 052313 (2007).

[8] A. S. Holevo, *IEEE Trans. Infor. Theory*, **44**, 269 (1998). M. V. Burnashev and A. S. Holevo, *Prob. of Inform. Trans.*, **34**, 97 (1998). B. Schumacher and M. D. Westmoreland, *Phys. Rev. A*, **56**, 131, (1997). M. Hayashi and H. Nagaoka: *IEEE Trans. Infor. Theory*, **49**, 1753 (2003).

T. Ogawa and H. Nagaoka, *IEEE Trans. Infor. Theory*, **53**, 2261 - 2266 (2007). A. Winter. *quant-ph/*0507045.
[9] A. Ambainis, *quant-ph/*9902053.
[10] M. Hayashi, A. Kawachi, and H. Kobayashi, *quant-ph/*0604174 *to appear in Quantum Information and Computation.*
[11] S. Virmani, M. Sacchi, M. B. Plenio and D. Markham, *Phys. Lett. A* **288**, 62 (2001)
[12] J. Walgate, A. J. Short, L. Hardy and V. Vedral, *Phys. Rev. Lett.* **85**, 4972 (2000)
[13] G. Alagic, C. Moore, and A. Russell, *quant-ph/*0511054, (2005); M. Grigni, L. J. Schulman, M. Vazirani, and U. Vazirani, *Combinatorica*, **24**, 137, (2004); S. Hallgren, *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, 653, (2002); S. Hallgren, *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, 468, (2005); J. Kempe and A. Shalev, *Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithms*, 1118, (2005); C. Moore, D. Rockmore, A. Russell, and L. J. Schulman. *Proceedings of the 15th ACM-SIAM Symposium on Discrete Algorithms*, 1106, (2004); C. Moore and A. Russell, *quant-ph/*0501177, (2005).
[14] Lehmann, E. L. 1986 *Testing statistical hypotheses*, Second edition, Wiley.
[15] M. Hayashi, D. Markham, M. Murao, M. Owari, and S. Virmani, *Phys. Rev. Lett.* **96**, 040501 (2006)
[16] E. M. Rains, *IEEE Trans. Info. Th.* **47**, 2921, (2001); M. J. Donald, M. Horodecki, O. Rudolph, *J. Math. Phys.* **43** 4252 (2002)
[17] K.M.R. Audenaert, J. Calsamiglia, Ll. Masanes, R. Munoz-Tapia, A. Acin, E. Bagan, and F. Verstraete, *Phys. Rev. Lett.* **98**, 160501 (2007).
[18] M. Nussbaum and A. Szkola, "A lower bound of Chernoff type for symmetric quantum hypothesis testing," quant-ph/0607216.
[19] S. Massar and S. Popescu, "Optimal Extraction of Information from Finite Quantum Ensembles," *Phys. Rev. Lett.*, **74**, 1259 (1995).
[20] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation*, (John Wiley & Sons, 1990).
[21] W. Hoeffding, "Asymptotically optimal test for multinomial distributions," *Ann. Math. Stat.*, **36**, 369-400 (1965).
[22] F. Hiai and D. Petz, "The proper formula for relative entropy and its asymptotics in quantum probability," *Comm. Math. Phys.*, **143**, 99–114, (1991).
[23] T. Ogawa and H. Nagaoka, "Strong converse and Stein's lemma in quantum hypothesis testing," *IEEE Trans. Infor. Theory*, **46**, 2428–2433 (2000); quant-ph/9906090 (1999).
[24] M. Hayashi, "Error Exponent in Asymmetric Quantum Hypothesis Testing and Its Application to Classical-Quantum Channel coding," quant-ph/0611013; to appear in *Phys. Rev. A.*
[25] H. Nagaoka, "The Converse Part of The Theorem for Quantum Hoeffding Bound," quant-ph/0611289.
[26] G. Vidal, R. Tarrach, *Phys. Rev. A* **59** (1999) 141-155
[27] C. H. Bennett, D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin, B. M. Terhal, *Phys.Rev.Lett.* **82**, 5385, (1999)
[28] private communication with S. Virmani and D. Markham
[29] M. Hayashi, D. Markham, M. Murao, M. Owari, and S. Virmani, arXiv:0710.1056. "*Entanglement and group symmetries: stabilizer, symmetric and antisymmetric states*".
[30] E. M. Rains, *IEEE Trans. Inf. Theory*, **47**, 2921 (2001); T. Hiroshima and M. Hayashi, *Phys. Rev. A*, **70**, 030302(R) (2004)
[31] W. K. Wootters and B. C. Fields, *Ann. Phys.*, **191**, 363 (1989).
[32] M. Owari and M. Hayashi, *in preparation.*

Appendix A. Proof of Eq.(20)

Now, we prove Eq.(20), which is used in proof of Theorem 2. Suppose $P \stackrel{\mathrm{def}}{=} \frac{1}{2}\sum_{i\neq j} |\overline{a_{ij}}\rangle\langle\overline{a_{ij}}| \otimes |\overline{b_{ij}}\rangle\langle\overline{b_{ij}}|$ and $Q \stackrel{\mathrm{def}}{=} \sum_{i\neq j}\{\sum_{k\neq i,j}\lambda_k + (\sqrt{\lambda_i}-\sqrt{\lambda_j})^2\}|e_i f_j\rangle\langle e_i f_j|$; that is, $\overline{T_0} = P + Q$. Then, by applying twirling operation over $U_{\overrightarrow{\theta}}$, we drive the following equality:

$$\begin{aligned}
&\int_0^{2\pi}\cdots\int_0^{2\pi} U_{\overrightarrow{\theta}} P U^\dagger_{\overrightarrow{\theta}} d\theta_1\cdots d\theta_d \\
&= (\sum_{j'=1}^{d} |e_{j'}\rangle\langle e_{j'}| \otimes |f_{j'}\rangle\langle f_{j'}|) P (\sum_{j'=1}^{d} |e_{j'}\rangle\langle e_{j'}| \otimes |f_{j'}\rangle\langle f_{j'}|) \\
&\quad + \sum_{i'\neq k'} (|e_{i'}\rangle\langle e_{i'}| \otimes |f_{k'}\rangle\langle f_{k'}|)\, P\, (|e_{i'}\rangle\langle e_{i'}| \otimes |f_{k'}\rangle\langle f_{k'}|)
\end{aligned}$$

This equality can be proven as follows: The action of a twirling operation (group-averaging) over a unitary representation of a compact topological group is equal to the action of the projection onto the subspace of all invariant elements under the group action [29]. For the action of $U_{\overrightarrow{\theta}}$ and $U^\dagger_{\overrightarrow{\theta}}$, the subspace (of operator-space $\mathfrak{B}(\mathcal{H})$) consisting of all the invariant element is spanned by the operators $\{|e_j f_k\rangle\langle e_j f_k|\}_{j\neq k}$ and $\{|e_j f_j\rangle\langle e_k f_k|\}_{ij}$. Therefore, we can easily see the above equation. For $i \neq j$, $i' \neq k'$, we have

$$\begin{aligned}
(\sum_{j'=1}^{d} |e_{j'}\rangle\langle e_{j'}| \otimes |f_{j'}\rangle\langle f_{j'}|)\, |\overline{a_{ij}}\rangle |\overline{b_{ij}}\rangle &= \sqrt{\lambda_j}|e_i f_i\rangle - \sqrt{\lambda_i}|e_j f_j\rangle \\
(|e_{i'}\rangle\langle e_{i'}| \otimes |f_{k'}\rangle\langle f_{k'}|)\, |\overline{a_{ij}}\rangle |\overline{b_{ij}}\rangle &= \delta_{i',i}\delta_{k',j}(\lambda_j\lambda_j)^{\frac{1}{4}}|e_i f_j\rangle - \delta_{i',j}\delta_{k',i}(\lambda_j\lambda_j)^{\frac{1}{4}}|e_j f_i\rangle .
\end{aligned}$$

Since

$$\begin{aligned}
&(\sqrt{\lambda_j}|e_i f_i\rangle - \sqrt{\lambda_i}|e_j f_j\rangle)(\sqrt{\lambda_j}\langle e_i f_i| - \sqrt{\lambda_i}\langle e_j f_j|) \\
&= \lambda_j |e_i f_i\rangle\langle e_i f_i| + \lambda_i |e_j f_j\rangle\langle e_j f_j| - \sqrt{\lambda_j}\sqrt{\lambda_i}|e_i f_i\rangle\langle e_j f_j| - \sqrt{\lambda_i}\sqrt{\lambda_j}|e_j f_j\rangle\langle e_i f_i| ,
\end{aligned}$$

we obtain

$$\begin{aligned}
&\int_0^{2\pi}\cdots\int_0^{2\pi} U_{\overrightarrow{\theta}} P U^\dagger_{\overrightarrow{\theta}} d\theta_1\cdots d\theta_d \\
&= \sum_{i\neq j}\Big(\lambda_j |e_i f_i\rangle\langle e_i f_i| + \lambda_i |e_j f_j\rangle\langle e_j f_j| - \sqrt{\lambda_j}\sqrt{\lambda_i}|e_i f_i\rangle\langle e_j f_j| - \sqrt{\lambda_i}\sqrt{\lambda_j}|e_j f_j\rangle\langle e_i f_i|\Big) \\
&\quad + \sum_{i\neq j}\sqrt{\lambda_i\lambda_j}|e_i f_j\rangle\langle e_i f_j| \\
&= (\sum_{i}^{d}|e_i f_i\rangle\langle e_i f_i|) - (\sum_{i=1}^{d}\sqrt{\lambda_i}|e_i f_i\rangle)(\sum_{i=1}^{d}\sqrt{\lambda_i}\langle e_i f_i|) + \sum_{i\neq j}\sqrt{\lambda_i\lambda_j}|e_i f_j\rangle\langle e_i f_j| .
\end{aligned}$$

In the same way, we can also show the equality $\int_0^{2\pi}\cdots\int_0^{2\pi} U_{\overrightarrow{\theta}} Q U^\dagger_{\overrightarrow{\theta}} d\theta_1\cdots d\theta_d = Q$; Q is invariant under the twirling operation.

Finally, we can calculate $\int_0^{2\pi}\cdots\int_0^{2\pi} U_{\overrightarrow{\theta}}\overline{T_0}U^{\dagger}_{\overrightarrow{\theta}}d\theta_1\cdots d\theta_d$ as follows:

$$\begin{aligned}
&\int_0^{2\pi}\cdots\int_0^{2\pi} U_{\overrightarrow{\theta}}\overline{T_0}U^{\dagger}_{\overrightarrow{\theta}}d\theta_1\cdots d\theta_d = \int_0^{2\pi}\cdots\int_0^{2\pi} U_{\overrightarrow{\theta}}(P+Q)U^{\dagger}_{\overrightarrow{\theta}}d\theta_1\cdots d\theta_d \\
&= (\sum_{i}^{d} |e_i f_i\rangle\langle e_i f_i|) - (\sum_{i=1}^{d}\sqrt{\lambda_i}\,|e_i f_i\rangle)(\sum_{i=1}^{d}\sqrt{\lambda_i}\,\langle e_i f_i|) + \sum_{i\neq j}\sqrt{\lambda_i\lambda_j}\,|e_i f_j\rangle\langle e_i f_j| \\
&\quad + \sum_{i\neq j}\{\sum_{k\neq i,j}\lambda_k + (\sqrt{\lambda_i}-\sqrt{\lambda_j})^2\}\,|e_i f_j\rangle\langle e_i f_j| \\
&= (\sum_{i}^{d} |e_i f_i\rangle\langle e_i f_i|) + (\sum_{i\neq j} |e_i f_j\rangle\langle e_i f_j|) - (\sum_{i=1}^{d}\sqrt{\lambda_i}\,|e_i f_i\rangle)(\sum_{i=1}^{d}\sqrt{\lambda_i}\,\langle e_i f_i|) \\
&\quad - \sum_{i\neq j}\sqrt{\lambda_i\lambda_j}\,|e_i f_j\rangle\langle e_i f_j| \\
&= I - T,
\end{aligned}$$

which proves Eq.(20). □